NIJMEGEN INSTITUTE
FOR COGNITION
AND INFORMATION

In search of expert knowledge

An exploratory study on expert knowledge in the domain of the classification of problem behaviors in children

Nicole Krol

nici

In search of expert knowledge

An exploratory study on expert knowledge in the domain
of the classification of problem behaviors in children

Een wetenschappelijke proeve op het gebied
van de Sociale Wetenschappen

Proefschrift
ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Nijmegen,
volgens besluit van het College van Decanen
in het openbaar te verdedigen
op woensdag 7 oktober 1992,
des namiddags te 1.30 precies

door

Nicole Petronella Clasina Maria Krol
geboren 28 november 1957
te Tilburg

NICI
Nijmegen Institute for Cognition
and Information, Nijmegen 1992

Promotor: Prof. Dr. E.E.J. De Bruyn
Co-promotor: Dr. J.H.L. van den Bercken

CIP-gegevens Koninklijke Bibliotheek, Den Haag

Krol, Nicole Petronella Clasina Maria

In Search Expert Knowledge: An exploratory study on expert knowledge in the domain f the classification of problem behaviors in children
Nicol Petronella Clasina Maria Krol
NIC , Nijmegen Institute for Cognition and Information
T esis Nijmegen
ISBN 90 373 0152 5 / CIP

Vormgeving: Ingrid Sewpersad met dank aan OptimaForma

In search of expert knowledge

An exploratory study on expert knowledge in the domain
of the classification of problem behaviors in children

Op zoek naar expertkennis

Een verkennende studie naar expertkennis in het
gebied van de classificatie van probleemgedragingen van kinderen

Acknowledgements

I wish to express my gratitude to all the people who contributed to this thesis in many different ways. In particular:

The diagnosticians and students for their willingness to participate in the research.

The Nederlandse Vereniging voor Pedagogen, Onderwijskundigen en Andragologen for giving permission for the use of their file of diagnostic reports.

Prof. Dr. F.C. Verhulst for providing the original factor scores for the empirical syndromes in the Dutch population.

Eric De Bruyn and John van den Bercken for their guidance and support.

Edward van Aarle for his guidance during the first part of the research project.

The members of the statistical consultancy group (RTD), especially Lex Bouts.

Anthony Jameson for correcting my English. All the remaining mistakes are mine, not his.

Ineke ten Cate, Elly Geeraarts, Trudeke Hoge, Bertine Mellink, Ingrid Sewpersad, Bas Sytstra, Marjolijn Vermande, Francoise Vermeulen for their assistance in collecting and analyzing the data.

My dear family, friends, and colleagues for their social and all the other sorts of support.

Contents

This thesis concerns a search for expert knowledge in the psychodiagnostic domain. In particular one important aspect of psychodiagnostics, the classification of problem behaviors, is addressed. The study is part of a research project that aims at the development of expert systems for the psychodiagnostic domain.

The referral of a child to a diagnostician is the start of a complex information-processing and problem-solving process where decisions are continually being made, implicitly as well as explicitly. The diagnostician has to find out what is going on, what the problem is exactly, and what can or has to be done. Research on clinical decision making has shown the fallibility of this decision process (e.g., Pijl, 1989; IJzendoorn & Bus, in press). In view of the large amount of information involved and the complexity of the task, this is not astonishing. As early as in 1954, Meehl proved the superiority of statistical formulae over clinical intuition and advocated the use of statistical formulae in personality assessment. Research trends favouring the actuarial method have exerted almost no influence upon clinical practice, and clinicians still prefer to use their heads rather than formulas (Dawes, Faust & Meehl, 1989; Kleinmuntz, 1990). The heuristics and biases described by Kahneman, Slovic and Tversky (1982) are referred to as causes of the poor performance in clinical decision making (Achenbach, 1985; Nurcombe & Fitzhenry-Coor, 1986). To improve diagnostic practice it seems necessary to develop (computerized) diagnostic decision support systems (De Bruyn, 1988). Especially the expert system as a decision aid is thought to offer promising possibilities for improving the quality of current psychodiagnostic practice (Ruyssenaars & De Bruyn, 1987).

Before one starts with the construction of diagnostic knowledge systems, it is important to determine the weaknesses and strengths of the clinical decision making process of the diagnostician in order to find out which parts of the process are in need of support. Thus, the study of clinical judgement, particularly the reasoning process of the expert, becomes an important research topic.

In the development of expert systems for psychodiagnostics, it seems important to evaluate the clinical knowledge used (Carroll, 1987). The descriptive modeling of clinical judgement, an approach that is often used in the development of expert systems, will otherwise preserve, reinforce, and perhaps even magnify, existing cognitive biases, since the judges being modelled may not be cognitively competent (Kleinmuntz, 1990). In many domains research on the value of expertise has focused on differences in knowledge characteristics and problem solving methods between persons with different levels of expertise (Boshuizen, 1989; Kleinmuntz, 1990; Kolodner, 1984; Kolodner & Simpson, 1986; Shanteau, 1988, in press). The added value that one in general assumes experience to provide, is not always demonstrated in studies. Therefore it can and should be called into question for the different tasks in the assessment process.

Following De Bruyn (1985, 1990), we view psychodiagnostics or clinical assessment as a cyclic process comprising four stages: (a) complaint analysis, (b) problem identification, (c) formulation and testing of hypotheses, and (d) treatment

formulation. This study focuses especially on the second component of this process, problem identification. During this stage, the diagnostician has to form a global representation of the problem on the basis of his or her knowledge and experience. He or she has to collect, select, combine and integrate information from different sources (parents, teachers, the child, etc.). A large amount of information has to be ordered or classified.

We first directed our study at the public knowledge available for this task, i.e. the sort of knowledge that can be found in textbooks and articles. Two important classification systems in the domain of child psychopathology were reviewed (DSM-III-R and the CBCL system). To determine whether the CBCL is a good candidate to be used as a data base in a (computerized) diagnostic knowledge system, we studied the representativeness of the CBCL system for the domain of problem behaviors of children in the age range of 6-11 years.

Next we addressed private psychodiagnostic classification knowledge, i.e.rules of thumb or heuristics the diagnostician actually uses when he or she has to solve a classification problem. Aiming to detect heuristic classification rules, we investigated the classification behavior of diagnosticians in two different classification tasks and validated the private classification knowledge. We also investigated the role of experience for this task. We compared experts' and novices' classification processes and the outcome of these processes. We also studied what classification structures develop in the minds of diagnosticians as a result of experience and how these structures relate to empirical classification structures.

The structure of the thesis is as follows: In Chapter 1 we present some issues related to the construction of expert systems. We address the notion of an expert, review research on clinical reasoning, focus on expert-novice differences, and address the role of experience.

In Chapter 2 we describe the psychodiagnostic process, discuss the importance of classification, and review some classification approaches. We conclude that the CBCL can be used as a data base for a diagnostic knowledge system.

The empirical studies are presented in Chapters 3, 4, and 5. In Chapter 3 we present a study on the representativeness of the CBCL system for the domain of problem behaviors of children in the age range of 6-11 years.

In Chapter 4 we describe an empirical study of the psychodiagnostic classification process in expert diagnosticians and we compare experts and novices with respect to both this process and the result of the process.

In Chapter 5 we address the classification structures as developed in the minds of diagnosticians as a consequence of experience and examine the empirical content of these structures.

Finally, in Chapter 6 we evaluate our research and discuss the conclusions and implications of the study for the development of expert systems, future research, and diagnostic practice.

Chapter 3 is a revised version of an article that has been published in the *Dutch Journal for Child Rearing and Education (Nederlands Tijdschrift voor Opvoeding, Vorming en Onderwijs)*. The empirical study on the content validity of the CBCL has also been reported as a letter to the editor in the *Journal of the American Academy of Child and Adolescent Psychiatry*. In both instances E.E.J. De Bruyn was the second author. Chapter 4 is an extended version of a paper that was presented at the SPUDM-13 conference in august 1991 and is accepted for publication in *Acta Psychologica*; E.E.J. De Bruyn and J.H.L. van den Bercken were co-authors of this paper. The extension involves mainly the Results section, in which the outcomes of the symptom classification task are reported.

1 Diagnostic expert knowledge

This chapter presents some issues related to the development of expert systems for the psychodiagnostic domain. We start with the vital question "What is an expert?" Then we review research on the reasoning processes of expert clinicians, focussing especially on differences between experts and novices and on the role of experience.

The terms *problem solving, diagnosis, decision making, reasoning, prediction,* and *judgement* all refer in one way or another to the task of the diagnostician and are used interchangeably in this chapter. So are the additions *intuitive, clinical,* and *subjective* in contrast to *statistical, mechanical,* and *actuarial,* all of which describe how the tasks are performed.

1.1 What is an expert?

Expert systems are knowledge-based computer programs that can solve real-life problems which require a considerable amount of expertise when solved by humans (Jackson, 1987). The knowledge that expert systems can include can be roughly divided into two types: public and private knowledge. Public knowledge includes the types of information found in textbooks or hand-books: definitions, facts and theories. Private knowledge consists of rules of thumb or heuristics that enable experts to make educated guesses when necessary.

The idea of constructing expert systems for a particular domain presupposes the existence of experts in that domain. But what is an expert? In artificial intelligence research, experts are those people who are considered to be the best at solving a particular problem (Shanteau, 1988). They have acquired this ability through the practice of the art, through experience. According to Kolodner, experts not only know more about their domain, but they have learned from experience how to apply and use that knowledge effectively (1984). Following this viewpoint, we distinguish two components in the use of the term expert in this thesis. We define an expert as someone (a) who has finished his or her professional education and is familiar with the public knowledge in the domain and (b) who has gained clinical experience and from this experience has learned how to apply and use that knowledge. In research it is this experience in particular that distinguishes experts from novices Shanteau (1988) and Dreyfus and Dreyfus (1986) present more detailed distinctions.

In our study we use the term *novice* in contrast with *expert* to refer to someone who has not finished his or her professional education and who has no clinical experience. The term expert refers to an experienced diagnostician or clinician.

1.2 Clinical reasoning

An expert system can be seen as a model of the reasoning process of an expert clinician in the form of a computer program. The construction of an expert system thus requires a detailed analysis of the problem-solving method or reasoning process of the expert. It is therefore not surprising that the idea of developing an expert system for the (psycho)diagnostic domain has revived interest in the reasoning processes of the clinician (Dowie & Elstein, 1988; Turk & Salovey, 1988).

As early as 1954, Meehl studied clinical reasoning - in particular how good clinicians are at making judgements - and started the debate about clinical or intuitive versus statistical prediction. Results from 20 empirical studies in which the actual predictive efficiencies of the two methods were compared showed that statistical prediction outperformed intuitive prediction. Only in one case was intuition better than statistics. Later, in 1965, Meehl was able to tally the *box-score* of 51 empirical investigations, of which 33 were judged to demonstrate the superiority of statistical over clinical data combination, 17 demonstrated 'approximate equality' of the two approaches, and only one study favored clinical over statistical prediction (see Wiggins, 1973). According to Meehl "it almost looks as if the first rule to follow in trying to predict the subsequent course of a patient's behavior is carefully to avoid talking to him, and that the second rule is to avoid thinking about him"! (1973, p. 64) In a noteworthy review Kleinmuntz stated that the main contribution of Meehl over the years has been to place judgement at center stage, to provide a sound rationale and empirical evidence for the scientific scrutiny of judgement (1990).

One main approach within the study of clinical reasoning is the problem-solving approach, which studies clinical reasoning from an information-processing standpoint (Elstein & Bordage, 1988). This information-processing view records and analyzes the steps and thoughts of the clinicians as they attempt to solve clinical problems. The goal is to describe the process associated with the particular task and to explain it in terms of basic psychological elements and principles.

A psychological principle basic to the understanding of clinical reasoning is the concept of *bounded* or *limited rationality* (cf. Newell and Simon, 1972). This principle emphasizes that limits exist to the human capacity for rational thought. In considering clinical reasoning, the most relevant limit is the relatively small capacity of working memory compared to the essentially infinite size of long-term memory. According to Elstein and Bordage this means that,

> *in a brief time, we cannot work efficiently with all we know about a problem or all the data that could be collected. Some common features of good and poor clinical reasoning are consequences of efforts to cope with this limitation. Given the limited size of working memory, one is literally required to process data serially, to select*

data carefully, to represent a clinical problem in simplified ways, and to work as rationally as possible within these simplified representations. (1988, p. 110)

The reasoning of chess players (De Groot, 1965), neurologists (Kleinmuntz, 1968), special educators (Bus & Kruizinga, 1989; IJzendoorn & Bus, in press), psychiatrists (Nurcombe & Fitzhenry-Coor, 1986) have been studied using this approach. Elstein, Shulman and Sprafka studied the reasoning of a group of experienced internists as they performed on a number of medical and non-medical problems (1978). They used thinking-aloud techniques and employed the problem solver's verbalizations as data. From these data they identified three units of analysis: cues, hypotheses and information search units. It was found that clinicians during the diagnostic process elicit salient cues from the patient. These cues are combined so as to delineate a clinical problem. Clinical problems are then resolved through a process of hypothesis generation and verification.

Elstein et al. (1978) also found that a small set of hypotheses are generated very early in the clinical encounter, when only limited data are available. The number of hypotheses rarely exceeded five. Comparable studies showed that clinicians generated an initial hypothesis in less than 1 minute after the first contact with the patient and that the average number of hypotheses was six. A limited set of alternative explanations seemed to simplify the search for data.

Elstein and Bordage remark that while principles used to simplify problems are often useful, they can nevertheless lead to certain errors, for example over-emphasizing positive findings or excessive data collection. Many heuristics and biases related to judgement under uncertainty are described in the volume edited by Kahneman, Slovic and Tversky (1982). Nurcombe and Fitzhenry-Coor (1986) illustrate some base rate errors in clinical psychiatry, for example the over-estimation of the probability that a disease is present as a result of previous experience (e.g., an encounter with a serious case), without adequately taking the base rates of the disease and or the finding into account. Also a phenomenon called *illusory correlation* can bias accurate judgement in clinical assessment (Achenbach, 1985; Kleinmuntz, 1990). This phenomenon was uncovered by Chapman and Chapman who reported that people, including clinical psychologists, tend to see a correlation between two events as being stronger than it actually is (cf. 1982). They showed that both clinicians and naive students saw massive illusory correlations between patients' traits on the one hand and their behavior on projective tests on the other hand. One example is the association between the drawing of atypical eyes with suspiciousness. They explained this finding by the same mechanism they discovered in an earlier study, i.e. the fact that subjects tend to see words with strong associative connections as occurring together in experimental presentations. In this earlier study people read word pairs (e.g., *bacon-tiger*, *bacon-eggs*) and were later asked about the word pairs. The subjects claimed that the pairs with a strong semantic association (e.g., *bacon-eggs*) occurred more often than the others (e.g., *bacon-tiger*), even though in fact every word-pair appeared as often as every other.

In view of the cognitive biases that influence the decision making processes of expert clinicians and the poor performance of experts in the domain of clinical psychology (Shanteau, 1988) it seems important to evaluate clinical knowledge before using it for expert system development. Now if we could determine an optimal response or answer to a diagnostic task, we could observe the extent to which expert responses deviate from the optimal responses and thereby evaluate the knowledge and clinical reasoning of the experts. Unfortunately, in clinical assessment and other so-called *poor performance domains* no reliable objective standards are available; this makes it difficult to define the optimal answer or response (Shanteau, 1988, in press). Therefore, in many domains research on the value of expertise has focused on differences in knowledge characteristics and problem solving between persons with different levels of experience.

1.3 Expert-novice differences

Research on expert-novice differences with respect to diagnostic reasoning in the psychiatric domain showed that experienced diagnosticians generate hypotheses earlier, are more consistent in evaluating them, and are more likely to provide rationales for inquiry plans. They also reach more accurate conclusions (Nurcombe & Fitzhenry-Coor, 1986). Kruizinga and Bus (1990) compared the diagnostic problem solving of advanced and beginning practitioners in the field of education. Their results contradict the expectation that advanced practitioners would work directly towards a conclusion. The advanced practitioners didn't hypothesize more in the beginning, nor did they seek information more efficiently later in the diagnostic process. In sum, this study did not confirm the hypothesis that advanced practitioners are more selective and flexible than beginners in solving problems. Reviewing medical expertise, Boshuizen (1989) reported studies that show that increasing knowledge and experience did not affect the time in which the first hypotheses were generated or the way in which they were verified or falsified.

Horowitz, Wright, Lowenstein, and Parad (1981a) examined the prototypes of three kinds of "problem child" formulated by experts or by returnees or novices. They found that the prototypes generated by returnees and novices were similar to those of experts but contained fewer features. The experts mentioned the largest number of prototypic features, while the novices mentioned the fewest.
Garb (1989) reviewed 55 studies which examined the effect of training and experience on judgement tasks relevant to clinical psychology. Neurologists who were reputed to be experts made more valid ratings than other judges, but expert psychologists making personality assessments were not more accurate than other judges. Experienced clinicians (not reputed to be experts) were never more accurate than less experienced clinicians. Similarly, clinicians almost never made more valid judgements than did graduate students.

Research findings certainly do not confirm the naive expectation of differences between experts and novices. Instead, research offers an inconsistent picture of the relationship between experience, problem solving, and expertise. These findings raise questions about the role of experience. Does experience make an expert?

1.4 The role of experience

Research discussed so far has not explained why some clinicians reach expert status while others with equal training and experience do not. Why doesn 't experience always provide expertise? The role of experience in the expert's reasoning processes has been studied by Kolodner (1984). Aiming to understand expertise in the domain of psychiatry, she examined the evolution from novice to expert. According to Kolodner and Simpson (1986), experience plays two important roles in problem solving. The first role of experience is to refine and modify the reasoning process. Successful experiences reinforce already known rules or previous hypotheses, whereas failures require analysis of the reasoning and knowledge that was used and modification of faulty rules and knowledge. The second role of experience is to provide exemplars upon which to base later decisions. Analogies to previous cases guide and focus later decision making. These two roles of experience are illustrated by the following example :

> **Dr. X sees a patient who shows classic signs of Major Depression. She has previously been diagnosed as Depressive and was treated in a mental hospital with antidepressants. She was sickly as a child, has had a drinking problem, and has had some unexplained physical illnesses. Dr. X concludes that she is suffering from Major Depression, Recurrent, without Melancholia and treats her with antidepressants. They seem to work, but the woman comes back complaining of additional major physical disorders.Taking a further history, the doctor finds that her unexplained medical problems have been numerous. Realizing that this is an important consideration, he makes a second diagnosis of Somatization Disorder.(Adapted from Kolodner & Simpson, 1986, p. 100.)**

Kolodner and Simpson report that Dr. X should learn from this case that it is important to consider medical history in choosing predominant clinical features and that Depression can camouflage Somatization Disorder. He should refine his rules for choosing predominant clinical features, i.e. modify his reasoning process (the first role of experience). When Dr. X sees a second patient diagnosed for Major Depression who also has unexplained medical problems, Kolodner and Simpson expect him to transfer his knowledge from the previous case to the new one and to consider whether the new patient might also have Somatization Disorder. This illustrates the second role of experience, providing exemplars.

An essential prerequisite for learning from experience is feedback (Kolodner, 1984; Shanteau, 1988). According to Garb (1989) one of the reasons why

experienced clinicians fail to make more valid judgements than less experienced clinicians is that they have trouble learning from experience. For some tasks feedback is unavailable; when available, it is often distorted because of biased cognitive processes like (a) inadequate hypothesis testing strategies, (b) the fallibility of memory, and (c) hindsight bias. Garb suggests that clinicians should be more willing to entertain alternative hypotheses and to revise their initial impressions of a patient. They should reduce their reliance on memory and instead keep good records and periodically review records. He urgues clinicians to rely more on their training (e.g. what they have read in the professional literature) and on decision-making aids (e.g. computer programs) than on their own experience.

1.5 Conclusion and Implication

Summarizing the different issues addressed in this chapter, we come to the conclusion that in the domain of psychodiagnostics it is difficult to define objective criteria to validate expert knowledge. Research on clinical reasoning has shown that the approach of modeling the clinical expert, an approach often used in expert system development, requires caution. The exploration of private knowledge in the domain of psychodiagnostics presupposes a validation of that knowledge. In two studies of this thesis (presented in chapters 4 and 5) we will explore private knowledge in the domain of psychodiagnostic classification and we will address the issue of the validation of this private knowledge.
But first we focus in the next two chapters on the public knowledge available in the domain of psychodiagnostic classification.

2 Psychodiagnostics and classification

This chapter addresses the public knowledge that is available in the domain of psychodiagnostic classification. Besides the place of classification in the psychodiagnostic process, we will focus on the purpose and principles of classification and discuss two important classification systems for child psychopathology.

2.1 Definition of terms: clinical assessment, psychodiagnostics, classification, and diagnosis

Clinical psychology still lacks a standardized vocabulary and many terms are not well-defined. We will therefore present our definition of some terms that will figure importantly in this chapter and the rest of this thesis.

Following Woody (1980), we define *clinical assessment* as a set of procedures for human services professionals, primarily designed to support idiographic problem solving. Clinical assessment is closely related to *psychodiagnostics*, which we define as a systematic decision procedure in which, simply stated, the problem of the client has to be examined and a decision has to be made as to what must be done (De Bruyn, 1985). So the terms *psychodiagnostics* and *clinical assessment* both refer to the conglomeration of activities that starts when someone is referred to a mental health service. In this thesis these two terms are used interchangeably.

Classification refers to any systematic grouping or ordering of features into groups or types. Classifying is the activity, i.e., of arranging features in classes or assigning them to categories. Different terms are used to describe the outcome of classifying, depending on the particlular scientific discipline involved: *group, type, category, class, syndrome, dysfunctional behavior cluster, disorder, or diagnosis.*

The terms diagnosis and classification are often used interchangeably in the literature. For instance, in the DSM-III-R classification system a diagnosis is made if a case meets the criteria of a category. A case may elicit, e.g., the diagnosis "conduct disorder" - i.e., the patient is classified as having a "conduct disorder".

We restrict the term *diagnosis* to referring to "a statement or conclusion concerning the nature or cause of some phenomenon" (Woolf, 1977, p.313). Diagnosing refers to the activity, the investigation or analysis of the cause or nature of the phenomenon.

Classifying and diagnosing both take place in psychodiagnostics. This thesis is concerned with psychodiagnostic classification, defined as the ordering or grouping of problematic behaviors (the phenomena) of the individual case into syndromes (the groups or types).

In the next section we analyze the role of classification in psychodiagnostics.

2.2 Classification in the psychodiagnostic process

In literature on psychodiagnostics, more and more attention has been focussed on the diagnostic process (De Bruyn, 1985, 1988, 1990; Rispens, Carlier & Schoorl, 1984; Van der Kooij & Knijff, 1986). De Bruyn (1985, 1990) has developed a normative framework for the diagnostic process. This model is based on the general methodological concept of the empirical cycle as formulated by de Groot (1961) and on the logico-normative approach to the diagnostic process as formulated by Westmeyer (1972). The structure of the diagnostic cycle is shown in Figure 2.1.

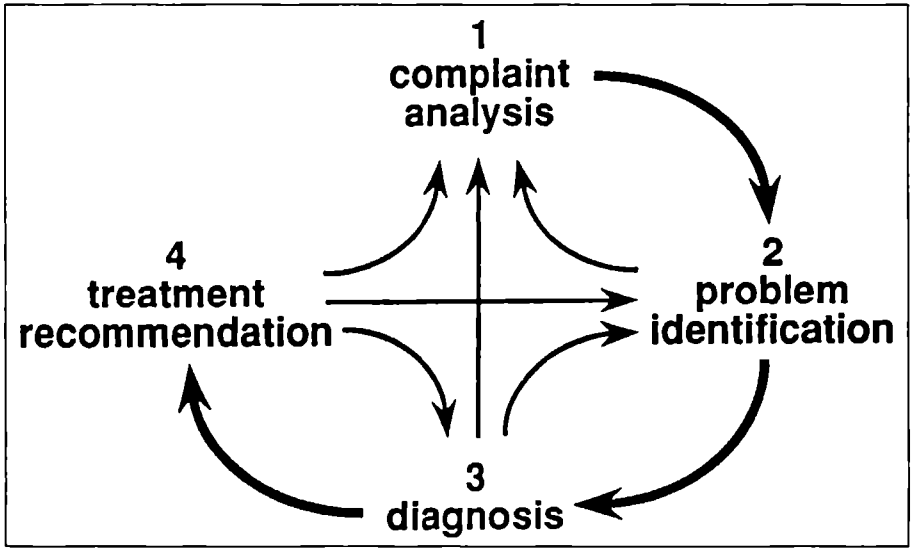

Figure 2.1: The structure of the diagnostic cycle

According to this framework, the diagnostic process starts with an analysis of the complaints of the client (1). The complaint analysis is followed by the problem identification (2). In this stage the task of the diagnostician is to sort the subjective complaints into empirically or theoretically based clusters of dysfunctional behavior. In the third stage, diagnosis, the clinician has to generate and test hypotheses about which conditions either elicit or sustain the identified problem. The fourth component is the recommendation of treatment. In this stage the diagnostician has to assess which criteria for therapy are applicable to his client and has to choose between alternative treatments by integrating the information gathered in the previous steps. If the diagnostician has not reached an acceptable level of certainty at the end of each of the components he or she can repeat some parts of the diagnostic sequence or even the entire sequence (De Bruyn, 1990).

We want to stress that this diagnostic cycle must be viewed as a prescriptive model and not as a representation of the diagnostic process in actual clinical practice. Research has shown that actual psychodiagnostic practice deviates from this model (Bus, 1989; Costello, 1988; De Bruyn, Pijnenburg & van Kessel, 1986).

In the above-mentioned model, classification takes place during the second phase of the diagnostic cycle problem identification. During this phase the diagnostician is confronted with all sorts of data from different sources such as parents, teachers, and the child. In order to process this large amount of information efficiently and to avoid an overwhelming cognitive overload, the diagnostician has to reduce and simplify the information. Achenbach, following Mischel, refers to this problem as one of *cognitive economics* (1985). The diagnostician thus has to select, combine and order the information. He or she has to bring structure into the description of the problem.

Despite the fact that classification as a cognitive process has received little attention in research on psychodiagnostics, classification systems have been developed to assist the diagnostician during this task. These systems provide categories into which the information can be ordered. While in medical practice classification systems are considered important and doctors are trained in using them, their usefulness is still questioned by clinicians in psychodiagnostic practice (Rispens, 1986). Clinicians are not always aware-or seem to forget- that they also use what Rutter has referred to as *private* classification systems. The criteria on which these *private* classifications are based remain vague and unclear, as does the impact of this classification on the generation of hypotheses. According to Rispens, a clinician cannot withstand the tendency to situate a client in the whole of his or her knowledge from experience, ordered in one way or other (1986). Achenbach reports that from the many features that characterize a case, clinicians abstract a few features and form a conceptual abstraction intended to capture important features of the case that link it to similar cases (1985).

Individuals and disorders can be classified in multiple ways, depending on the attributes and principles chosen as a basis for classification. According to Achenbach (1985, 1988), *taxonomy* involves the grouping of cases according to their distinctive features. Classification comprises groupings based on extrinsic criteria as well as those based on features of the cases themselves. For example, "a clinic may assign cases to therapists who have therapy hours when the cases are referred. Thereafter, the clinic may classify each case according to the child's therapist. This system may be convenient for billing, record keeping, and assignment of clinical responsibilities, but it does not reflect intrinsic differences among the cases" (Achenbach, 1985, p.152).

According to Achenbach, each category of a taxonomy represents a hypothetical construct defined by attributes that are singled out from all the other attributes that characterize individuals.

We use the terms *classification system* and *taxonomy* interchangeably.

2.3 Methodological principles of classification systems

The basic principles of classification systems for child psychopathology, along with their purposes and requirements, have been reported repeatedly in literature (Achenbach, 1985, 1988; Blashfield, 1984; Rispens, 1986; Rutter, 1975; Rutter & Gould, 1985; Rutter, Tuma & Lann, 1988). The general purpose of classification systems is to provide "a kind of language-an agreed set of terms-by which clinicians and research workers can describe the disorders they investigate and treat. Only if there is uniformity in the usage of descriptive and diagnostic terms can meaning be attached to clinical reports, research findings or hospital statistics" (Rutter and Gould, 1985, p.304). In the diagnostic process the use of classification systems is recommended as aids in mapping the relations between the diverse behaviorial aspects of the problem and in describing the problem behavior.

Rutter and Gould have proposed basic requirements for an adequate classification system for child psychopathology, including the following: The classification system should be based on facts, not on theories or concepts, and it should be defined in operational terms. The terms must be used in the same way by different clinicians, in other words, the system should be reliable. Furthermore, the system should classify disorders or problems, as opposed to children. To illustrate this criterion, one often makes a comparison with the medical classification of disorders, where we do not talk about a measly child but about a child who *has* the measles (measles being the disorder). It is better not to talk about a neurotic child but about a child who has a neurotic disorder. (Note that the DSM-III-R manual discourages the use of expressions such as "a schizophrenic" and suggests using an expression like "a person with Schizophrenia" instead).The classification must also take into account the fact that the child is a developing organism. The same abnormality can manifest itself in different ways at different ages. The system should provide adequate coverage, so that important disorders are not omitted. The differentiations between categories should be valid and should have meaning in terms of aetiology, symptomatology, course, response to treatment, or other variables.

Rutter and Gould conclude that at the moment there is no really satisfactory classification system for child psychopathology.

2.4 Approaches to classification

In the field of child psychopathology, different taxonomic or classification systems have been developed. Both Blashfield (1984) and Achenbach (1985, 1988) have extensively discussed the leading taxonomic approaches, such as the Kraepelinian and the multivariate approach and the classification systems related to it. If one looks at the impact of these systems, two systems seem to be most

important for the domain of child psychopathology, i.e. the DSM-III-R system and the CBCL system. In a preliminary stage of the present research project, Traudes (1987) discussed and compared the different approaches in view of the question of which system was the best candidate for implementation in a diagnostic knowledge system. He concluded that the CBCL system was the best suited. Consequently it was decided to concentrate further research on the CBCL system. To document this decision, we describe both DSM-III-R and CBCL and compare them with respect to some general issues. For a more detailed discussion and comparison, we refer to Traudes (1987).

2.4.1 A Kraepelinian taxonomic approach: DSM-III-R

The Kraepelinian taxonomic approach arose through the application of the nineteenth-century medical nosology to mental disorders. It was based on the assumption that clinical descriptions would permit the identification of syndromes for which organic causes would eventually be found. All current versions of Kraepelinian taxonomies of childhood disorders (e.g., DSM) have based their categories on clinical constructs for which descriptions were formulated by processes of discussion and negotiation, i.e. on so-called *armchair theories* according to Rutter.

The DSM taxonomy follows the Kraepelinian approach of categorical nosology. The most recent versions of the DSM no longer include aetiological theories. Since the first edition in 1952, the Diagnostic Statistical Manual (DSM) has been revised repeatedly. The latest version, the DSM-III-R, was published in 1987.

In DSM-III-R a mental disorder is conceptualized as

> *a clinically significant behaviorial or psychological syndrome or pattern that occurs in a person and that is associated with present distress (a painful symptom) or disability (impairment in one or more important areas of functioning) or with a significantly increased risk of suffering death, pain, disability, or an important loss of freedom. In addition, this syndrome or pattern must not be merely an expectable response to a particular event, e.g., the death of a loved one. Whatever its original cause, it must currently be considered a manifestation or a behaviorial, psychological, or biological dysfunction in the person. (APA, 1987, introduction, xxii.)*

The DSM-III-R system is a multiaxial system and consists of five so-called axes. The first three axes constitute the official diagnostic assessment. Axis I contains clinical syndromes and the V codes; the latter represent behaviorial or psychological problems that may appropriately be a focus for professional attention or treatment even though these are not attributable to a mental disorder. Axis II contains personality and developmental disorders. Axis III lists physical illness, Axis IV provides ratings of psychosocial stress, and Axis V provides ratings of the person's highest level of adaptive functioning during the previous year. It contains

a special section concerning childhood disorders.

The definitions of the disorders are limited to descriptions of the clinical features of the disorders. The text of the DSM-III-R describes each disorder in terms of current knowledge in the following areas: essential features, associated features, age at onset, course, impairment, complications, predisposing factors, prevalence, sex ratio, familial pattern, and differential diagnosis.

The DSM-III-R provides specific diagnostic criteria and rules for each disorder. Most childhood disorders contain three or four types of criteria (not all criteria are always applied to the disorders): (a) a list of descriptive features of which a specified number must be present; (b) a minimum period during which the behaviors must be present, e.g. 6 months; (c) an age criterion, e.g. onset before age of seven; (d) criteria for excluding a particular disorder if it is "due to" certain other disorders. These criteria and rules are meant to improve the reliability of the classification process and are based on clinical judgement.

An example of the diagnostic criteria for Oppositional Defiant Disorder is given below:

Diagnostic criteria for 313.81 Oppositional Defiant Disorder

Note: **Consider a criterion met only if the behavior is considerable more frequent than that of most people of the same mental age.**

A. A disturbance of at least six months during which at least five of the following are present:

(1) often loses temper

(2) often argues with adults

(3) often actively defies or refuses adult requests or rules, e.g., refuses to do chores at home

(4) often deliberately does things that annoy other people, e.g., grabs other childeren's hat

(5) often blames others for his or her own mistakes

(6) is often touchy or easily annoyed by others

(7) is often angry and resentful

(8) is often spiteful or vindictive

(9) often swears or uses obscene language

Note: **The above items are listed in descending order of discriminating power based on data from a national field trial of the DSM-III-R criteria for Disruptive Behavior Disorders.**

B. Does not meet the criteria for Conduct Disorder, and does not occur exclusively during the course of a psychotic disorder, Dysthymia, or a Major Depressive, Hypomanic, or Manic Episode.

Criteria for severity of Oppositional Defiant Disorder

Mild: **Few, if, any symptoms in excess of those required to make the diagnosis *and* only minimal or no impairment in school and social functioning.**

Moderate: **Symptoms or functional impairment intermediate between "mild" and "severe"**

Severe: **Many symptoms in excess of those required to make the diagnosis *and* significant and pervasive impairment in functioning at home and school and with other adults and peers**

(Adapted from the DSM-III-R manual, p.57.)

2.4.2 *The multivariate taxonomic approach: The CBCL system (empirically based assessment)*

The multivariate approach uses statistical procedures to identify taxonomic groupings of features or attributes. Multivariate analyses are used to measure the actual covariation between attributes. Subjective judgement is involved in selecting the samples and attributes to be analyzed, the analytic methods, and the mathematical criteria. Multivariate studies have shown considerable similarity among independently derived syndromes of behavior problems in child psychopathology despite differences in rating instruments, subject samples and analytic methods (for a review of this work see Achenbach & Edelbrock, 1978).

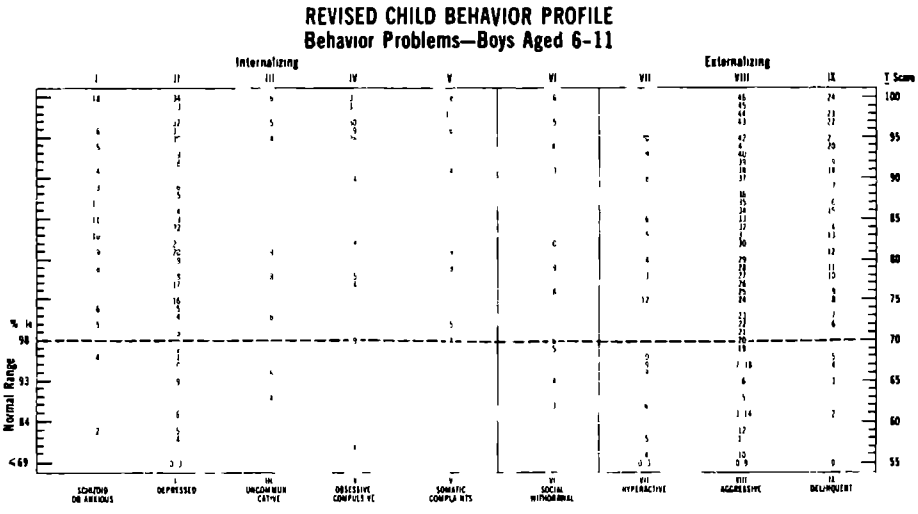

Figure 2.2: Example of the child behavior profile for boys aged 6-11 (reprinted from the CBCL manual, 1983, p. 21).

Achenbach and Edelbrock have developed a classification system based on multivariate findings, the *empirically based assessment* approach. This approach followed

> *psychometric principles, including the use of standardized procedures, multiple aggregated items, normative-developmental reference groups, and establishment of reliability and validity. This approach does not dictate nor preclude particular theoretical explanations for the phenomena to be assessed. It makes use of multiple sources of data to avoid the limitations and biases affecting each source of data taken alone (Achenbach & McConaughy, 1987, p.16).*

Achenbach in fact started with this approach in 1966 by developing a standardized instrument to rate children's problem behaviors. This resulted in the Child Behavior Checklist, a standardized format to record the behaviorial problems and competencies of children as reported by the parents.

Syndromes of behavior problems were identified through principal com-

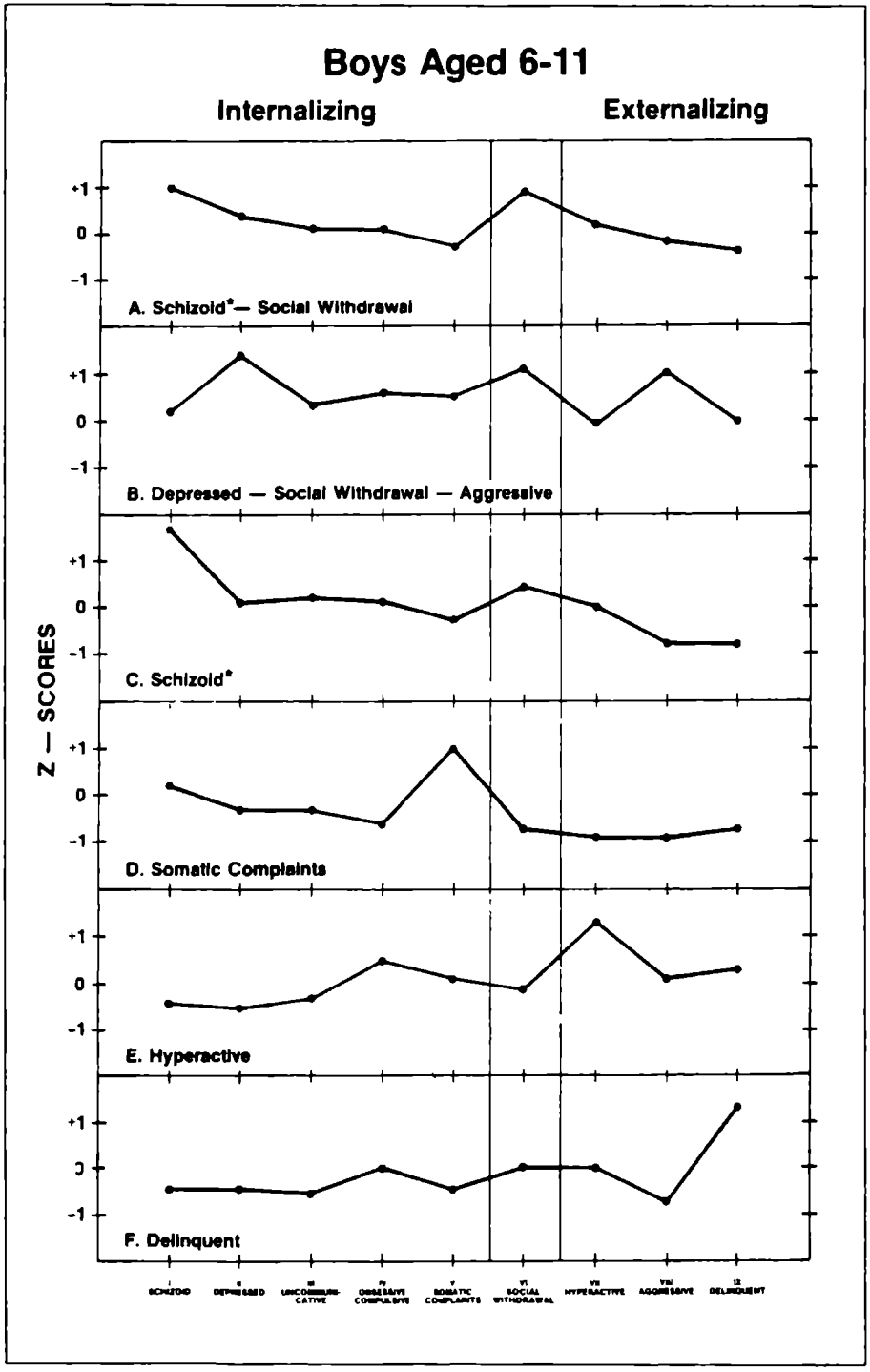

Figure 2.3: Profile pattern types found for boys aged 6-11 (reproduced from the CBCL manual, 1983, p. 78

ponents analyses of the CBCL behavior problems rated by parents of children referred to outpatient mental health services (for detailed information on different statistical analyses performed, see the CBCL manual, 1983). Separate analyses were performed for children of each sex at ages 4 through 5, 6 through 11, and 12 through 16. For each sex/age group, eight or nine behavior problem scales were found, and these were given descriptive labels. To provide a comprehensive view of the behavior problems, these scales are combined in the Child Behavior Profile (see figure 2.2). Using CBCL ratings from parents of randomly selected children who had not received mental health services for at least the preceding year, the behavior problem scales were standardized.

Cluster analyses of the profiles of clinically referred children have identified either six or seven profile types, replicated across two randomly selected clinical samples (see figure 2.3).

The cluster analyses revealed profile patterns with peaks on several syndrome scales. For instance, for boys aged 6 through 11 a profile pattern showed peaks on depression, social withdrawal and aggression: Profile patterns with a peak on a single syndrome scale, such as hyperactivity, were also found.

Verhulst adapted the CBCL system for the Dutch population. He replicated both the empirical grouping of problem behaviors in a Dutch clinical population and the standardization of the behavior problem scales (Koot & Verhulst, 1990; Verhulst, 1985; Verhulst, Koot, Akkerhuis & Veerman, 1990).

2.5 Some general issues

2.5.1 Categorical versus dimensional systems

A question often raised in the classification literature is whether the classification scheme should be categorical or dimensional. In a categorical classification system, symptoms are assigned to one or more categories, e.g., a child is classified as having a depressive disorder. The DSM-III-R system is often given as example of the categorical classification approach, one whose categories are not mutually exclusive, in that the symptoms of a subject can be assigned to more than one category (Rispens, 1986). The ICD-9 is based in the assumption that in most cases there is just one syndrome (Rutter & Tuma, 1988).

By contrast, the dimensional classification model represents the symptoms of subjects in terms of all of the dimensional continua included in the model. An example of this approach is Eysenck's three-dimensional model of psychopathology, which comprises the dimensions psychoticism, neuroticism, and introversion-extraversion (see, e.g., Eysenck, 1960). Eysenck's factor-analytic studies of personality with psychiatric patients showed, e.g., that hysterics were not only extremely high on the extraversion dimension: they were also high on neuroticism (Blashfield, 1984). The Child Behavior Profile of the CBCL system

also reflects the dimensional approach. For instance, a child who scores high on the depression dimension might also score high on the aggressive dimension and low on the hyperactive dimension.

The fact that in the dimensional approach more of the acquired information is retained is considered an important positive feature. Another advantage of dimensional models is that they are generally tied to sophisticated measurement systems for psychopathology. The categorical approach is often associated with traditional psychiatric classification and consequently also with its drawbacks. Few attempts have been made however, to explicate the categorical model; consequently the class of categorical models includes a number of possible structural variations (Blashfield, 1984). According to Blashfield, it is futile to decide which model is superior without a better understanding of categorical models. Rutter and Tuma (1988) believe that the most appropriate classification scheme is likely to be a mix of the categorical and dimensional approaches, rather than an pure form of one or the other. They point out that for some disorders the dimensional approach doesn't apply, e.g., developmental language disorder and nocturnal enuresis.

2.5.2 Categories versus prototypes

The distinction between categories and prototypes is related to the issue of category membership in the classification system. In a classification system based on categories such as those embodied in the DSM-III-R, a clinician decides whether or not the subject meets the criteria for a disorder. Category membership is viewed as an "all-or-none" decision. This approach corresponds with what may be called the "classical" view in the study of concepts and categories. This classical view holds that something is a member of a particular category because it satisfies the set of necessary and sufficient conditions which constitute the category's defining properties (McCauley, 1987, p.289). For example an object is categorized as a chair if it exhibits a certain set of defining features, e.g., four legs, a seat, a back, and arms.

The work of Rosch, Mervis, Gray, Johnson and Boyes-Bream (1976) demonstrated experimentally that natural categories have a graded structure, i.e. include more or less typical exemplars. Within a given category, some members are usually regarded as especially good or typical examples; the very best are called "prototypes". Other members of the category are relatively more marginal, and the category boundary is often poorly defined. For example the prototypical chair is the four-legged straightbacked kind often seen in a dining room. Modernistic single-pedestal armchairs are much less typical of that category (example adapted from Neisser, 1987). According to Neisser this "graded structure" appears in every kind of category.

The problems Rosch et al. (1976) described in conforming object categorization to the classical view are also recognized for psychiatric categorization:

there are no agreed-upon defining features for diagnostic categories, and the existence of borderline cases is assumed (Cantor, Smith, French, & Mezzich, 1980). By asking 13 experienced clinicians to list "clinical features that characterize the prototypical patient", Cantor et al. (1980) demonstrated that the prototypical view of categorization is in correspondence with the way clinicians actually think about and use diagnostic categories. According to Achenbach there is little reason to assume that most childhood disorders actually exist in categorical form. He suggests viewing disorders in terms of quantifiable prototypic syndromes, thereby avoiding forced choices between categories and diminishing the risk of neglecting features that do not conform to the most salient category (1988). The empirically based assessment method (i.e., the CBCL) represents this prototypical view of categorization. It consists of empirical prototypes derived through multivariate analyses of features scored in samples of individuals containing features that co-occur together. Conceptual prototypes contain features people think of as occurring together.

The prototypical view of categorization can also be applied in a categorical classification system. The use of severity criteria for some disorders in the DSM-III-R system represents a minor adaptation to a more prototypical view of categorization.

2.5.3 Reliability

In discussing the reliability of the Kraepelinian approaches such as the DSM, Achenbach refers to the fact that most studies have focused on agreement between judges in categorizing cases in which the judges obtained identical assessment data. The reliability of the assessment data itself is neglected in Kraepelinian taxonomies. For instance the DSM lacks standardized procedures for obtaining the necessary assessment data. The lack of an empirical data base for DSM categories of childhood disorders may contribute to the poor reliability often found for them, according to Achenbach (1989). Although the DISC (Diagnostic Interview Schedule for Children) was developed to assess specific DSM-III criteria via structured interviews with children, the test-retest reliability of the children's responses was poor (Edelbrock & Costello, 1988).

Most multivariate taxonomic approaches employ standardized assessment procedures, such as checklist ratings. Numerous studies have reported reliability and stability of scores for multivariate syndromes (Achenbach and Edelbrock, 1983; Achenbach, 1985: Verhulst et al., 1990).

2.5.4 Validity

The lack of substantive knowledge about childhood disorders makes it difficult to validate the categories of the classification systems. The categories or groupings in the DSM-III-R represent the disorders that a group of experts have thought of.

They were not derived from data on actual samples of children. The defining features and cutoff points for DSM diagnostic categories are not based on demonstrated discrimination between a clinical and a nonclinical sample, and the DSM criteria have not been empirically calibrated for different ages. According to Rutter and Gould, for many child psychiatric disorders in DSM III research evidence is lacking, with the consequence that "criteria have had to be pulled out of the air for conditions not adequately validated" (1985, p. 316). In a preliminary report of the DSM-IV work group, it is recognized that revisions of the categories should be mostly based on empirical data and that "unsupported 'expert' opinion" will not be sufficient (1990). The relation between empirically derived syndromes and the DSM-III syndromes has been studied, and some DSM-III syndromes correspond to CBCL syndromes (see Achenbach, 1985; Vermande, 1991).

A classification scheme is often validated by relating it to other variables, such as course or response to treatment. For most DSM categories the diagnostic criteria have not yet been fully validated with respect to important correlates such as clinical course, outcome, family history and treatment response (APA, 1987).

The validity of the empirically derived syndromes has been demonstrated by studies that showed significant correlations between syndrome descriptions made using different instruments (Achenbach & Edelbrock, 1978; Achenbach, Conners, Quay, Verhulst, & Howell, 1989; Rutter & Gould, 1985). The fact that multivariate studies of the CBCL in different countries obtained similar groupings of features supports the validity of these groupings (Achenbach, Althaus, Baron, & Verhulst, 1987; Verhulst, Achenbach, Althaus, & Akkerhuis, 1988). Discriminative validity of the empirically based syndromes has also been reported (Achenbach, 1985; Achenbach & Edelbrock, 1983; Verhulst, 1985; Verhulst et al., 1990).

2.6 Conclusion

Evaluating the two classification systems in view of our ultimate goal, the construction of a diagnostic knowledge based system, we come to the conclusion that the empirically based assessment approach is the most solid one to be used as a database for such a system. Although the DSM-III classification system can be called an "expert" system in the sense that it is based on the judgements of prominent clinicians, its has some important shortcomings. The systems lacks well-validated operational definitions for childhood disorders, has no standardized assessment instrument, and needs a better empirical basis for the grouping of the features. This diminishes the expert status of the system as a classification scheme for childhood disorders.

Despite the favourable judgement of the empirically based assessment method, we considered it appropriate to investigate the item base of this method

for our domain in more detail. The reason for this investigation and the presentation of the empirical study will be found in the next chapter.

3 The content validity of the CBCL

3.1 Introduction

One of the necessary components of a computerized diagnostic knowledge system to support the assessment of behavior problems in children is an empirically based, standardized, and reliable classification system for behavior disorders. As argued in chapter 2, the assessment procedure developed by Achenbach and Edelbrock (1983) and applied in the Netherlands by Verhulst (1985) seems to be a good candidate for implementation

But using a standardized classification system as a fixed component of a computerized diagnostic knowledge system involves certain risks. A comprehensive diagnostic knowledge system can suggest, albeit unintentionally, that the entire domain of clinical problems has been captured. If certain problem behaviors are not represented in the classification system, (in this case the CBCL), they will not be detected in any case to which the system is applied. As regards the quality of the classification system as a declarative data base, one of the simplest and most primary concerns is that of its representativeness of the domain, or, in other words, its content validity. According to Farrell (1984), this type of validity is perhaps one of the most important desiderata for computerized assessment systems.

According to Drenth (1975), content validity is a disputed concept in the testing literature and is difficult to verify empirically. It is often determined through a rational analysis of the content of a test, on the basis of individual, subjective judgement, it is therefore more subject to error than other types of validity. We refer to content validity in its most strict sense, namely "the extent to which the content of a test is a good representation of the behavior domain for which the score of a test is intended" (De Bruyn, 1985, p.181; translation by the present author). Content validity depends on the adequacy of the representation of the domain. It has been repeatedly stated in the literature (Allen & Yen, 1979; De Bruyn, 1985; Drenth, 1975) that it is difficult to determine this type of validity. This would require an exhaustive enumeration of the domain of behaviors to be measured. Only then can one check whether the items of the test reflect the total domain of behaviors adequately.

If the CBCL is to be used as declarative data base in a computerized diagnostic knowledge system for the assessment of problem behaviors, it is supposed to reflect the domain of problem behaviors for which children are referred to mental health services. In view of this purpose, it is necessary to examine the representativeness of the CBCL items for this domain. We will do this by reviewing the development of the item pool of the CBCL and by presenting and discussing an empirical study. We examined only the 118 behavior problem items of the CBCL, because our focus is on symptoms reflecting problem behaviors. The social competence items of the CBCL were not examined. For practical reason we limited our domain to problem behaviors of children aged 6-11.

3.1.1 Development of the item pool of the CBCL

Content validity depends on the care with which a domain of content has been represented in the test and on the scope of the sample, i.e. whether the items broadly sample the domain. Achenbach reports (1966, 1978a, 1978b) that the development of the behavior problem items of the CBCL was based on a symptom checklist he used in a factor-analytic study of case history data. To collect data for an empirical grouping of attributes, he used a procedure in which he attempted to minimize the influence of systematic biases in the observer. He sought "symptom categories which were objective, and required as little inference as possible" (1966, p.7). He constructed a checklist from items which "regularly appeared in previous studies which seemed to involve minimal inference, which could be considered mutually exclusive with regard to specific observations, and which were not excessively molecular" (1966, p.7). To obtain more symptom categories, 40 case histories of a child psychiatry unit of the university were analyzed. Finally, a list of 91 symptoms was constructed. "Symptom" referred to deviant behaviors, postures, attitudes, or verbalizations accepted as reasons for psychiatric concern. This original checklist was adapted to be used by parents by simplifying the words, by expanding the response alternatives, and by adding new items in consultation with clinicians. Pilot editions were further revised on the basis of item analysis and feedback from parents, clinicians and paraprofessionals (1978a). In this way the final problem behavior list, the Child Behavior Checklist (CBCL), was constructed, containing 118 problem behaviors.

In view of the large number of case histories used, (1,000 child psychiatric patients, 1978a), one might suppose that the behavior problem items of the CBCL adequately represent the aforementioned domain. But as already stated, it is difficult to check for this type of validity. In the manual of the CBCL its content validity is justified through reference to the fact that, on nearly all of the items in the list, the clinically-referred children received higher scores than the non-referred children (1983). This fact however points to the discriminative validity of the CBCL, which has also been demonstrated for Dutch children by Verhulst (1985). The representativeness of the item pool of the CBCL for a Dutch clinical population has to our knowledge not been checked.

Achenbach and Edelbrock themselves stress the fact that "prospective users should judge whether the content of the CBCL is appropriate for their particular purposes"(1983, p.51). In view of our purpose (the design of a computerized diagnostic knowledge system), we cannot fail to check the elementary aspect of the content validity of the CBCL. We decided to investigate whether the item sample of the CBCL broadly covers the domain of problem behaviors due to which children aged 6-11 are referred in Dutch clinical practice.
Two studies addressing this issue are reported in the present chapter.

Study 1: The content validity of the standard CBCL
In this study we attempted to assess whether the items of the CBCL appropriately represent problem behaviors of children aged 6 to 11 as reported in clinical practice. Following Achenbach (1966) we collected descriptions of problem behaviors from case histories. Subsequently, we compared these descriptions with the items of the CBCL.

Study 2: The content validity of an expanded CBCL
The results of Study 1 led to an expansion of the CBCL and to a study with this expanded behavior checklist in a practical diagnostic situation. In this study it was examined whether the problem behaviors that appear in a specific clinical group, namely children with motor dysfunctions, can be determined with the expanded behavior checklist. Futhermore, the added value of the CBCL method above an open-question during the phase of problem identification method was analyzed (Ten Cate & Hoge, 1989).

3.2 Study 1: The content validity of the standard CBCL

3.2.1 Method

Sampling procedure
The population consisted of case histories of children who had been clinically referred because of behavior problems. A sample was drawn from a file of standardized diagnostic reports of a Dutch professional organization for diagnosticians, the NVO. Diagnosticians who want to obtain a professional registration within this organization have to submit five completely diagnosed case histories. As a result, this organization has at its disposal a file consisting of hundreds of standardized reports obtained from several different mental health services. One of the standard requirements for the reports submitted is that the diagnostician has to give a clear and complete description of the child's problems. This greatly facilitated the collecting of the descriptions of problem behaviors in the present study. From all of the reports submitted by the 128 diagnosticians registered between 1980 and 1987, 582 case histories formed a pool from which the sample to be used in the present study was drawn.

The sample was obtained as follows: First, the original file was stratified according to the different mental health services involved. Reports submitted by diagnosticians who worked in settings that were not primarily concerned with treating problem behaviors of intellectually normal children were disregarded; similarly reports in which the diagnostician's place of work was not mentioned were omitted. For the remaining cases, information was collected about the child's age and sex.These steps yielded a file consisting of 324 case histories, involving 97 girls and 227 boys, aged 6-11. In Table 3.1. the distribution of the

diagnosticians over different mental health services and the distribution of boys and girls over the case histories is given.

Finally, a sample was taken from this file conforming to the gender distribution and to the distribution of mental health services. It consisted of 93 case histories, involving 26 girls and 67 boys, all aged 6 to 11.

Table 3.1
Overview of cases from which the sample for Study 1 was drawn.

		Diagnosticians	Girls*	Boys**
Ambulant services				
1.	RIAGG/MOB	20	15.4 % (15)	16.7 % (38)
2.	Team VroegTijdige Onderkenning	1	0.0 % (0)	0.4 % (1)
3.	GG en GD	3	3.0 % (3)	3.8 % (9)
4.	OBD/SBD/SAD	30	38.1 % (37)	35.6 % (81)
5.	Ambulatorium Orthopedagogiek	6	5.1 % (5)	6.6 % (15)
6.	PAO-Cursus Pedodiagnostiek	13	9.2 % (9)	12.7 % (29)
7.	Jeugdhuis en adviesburo	1	1.0 % (1)	0.4 % (1)
8.	Buro voor Pedagogische hulp	1	2.0 % (2)	0.8 % (2)
9.	Stotterproject	1	1.0 % (1)	1.7 % (4)
10.	Privé praktijk	2	2.0 % (2)	0.0 % (0)
11.	School-beroepskeuze buro**	1		
(Semi)-residential services				
12.	Dagcentrum voor schoolgaande jeugd	1	0.0 % (0)	1.7 % (4)
13.	Medisch kindertehuis	2	4.1 % (4)	0.8 % (2)
14.	Medisch kleuterdagverblijf	3	0.0 % (0)	0.8 % (2)
15.	(Kinder)ziekenhuis(5) Epilepsiecentrum(1)	6	3.0 % (3)	5.2 % (12)
16.	(Poli)kliniek kinderpsychiatrie	7	6.1 % (6)	2.6 % (6)
17.	Behandelingstehuis(5) Kindertehuis(1)	6	7.2 % (7)	6.1 % (14)
18.	Pedologisch instituut	4	2.0 % (2)	3.0 % (7)
19.	Behandelingstehuis voor visueel gehandicapten**	4		
20.	(Semi)residentiele voorziening voor geestelijk gehandicaptenb	3		
21.	Schoolinternaat voor licahmelijk gehandicapten jongeren**	1		
Special schools**				
21.	LOM/MLK/ZMLK	5		
22.	Internaatsschool	1		
Worksetting not mentioned**		12		
Total		134***	100 % (97)	100 % (227)

* Each percentage is relative to the total number of cases for the column. Absolute numbers are given in parentheses.
** These mental health services were omitted from the file from which the sample was taken.
*** There were 128 diagnosticians involved, 6 of whom had worked for 2 different services.

Data base definition

From the 93 case histories, the problem descriptions were extracted. In the reports the problem behaviors were described under different headings, such as problem description, complaint analysis, or analysis of problem area. Therefore all text in

the report preceding the formulation of one or more hypotheses (the next phase of the diagnostic process) was analyzed. Only problem behaviors of the child itself were taken into account; problems that other persons had with the child (e.g., the mother panics if child doesn't come home on time) were disregarded. The description was copied literally from the report; no interpretation or weighting was allowed. For example, if the problem description characterized the child as "naughty", this was taken literally as the problem description.

Subsequently, we compared these problem behaviors with the items of the CBCL by writing the corresponding item number of the CBCL (the Dutch version, adapted by Verhulst) behind each of the collected problem behaviors as shown in the following example:

Description of problem behavior	CBCL itemnumber
Is very nervous	45 (= nervous, high-strung, or tense)
Panics easily if he cannot succeed	
Weak concentration	8 (= can't concentrate, can't pay attention for long)

No special instruction was given as to whether the two characterizations had to be literally, or semantically equivalent.

Problem behaviors that could not be matched to one of the CBCL items were examined more closely. Problem behaviors considered to involve specific learning problems (e.g. "has difficulty with visual analysis and synthesis") or problem behaviors clearly related to school situations were omitted, because the CBCL was not developed to assess these kinds of problem behaviors. On the base of semantic equivalence we combined the remaining descriptions. For example the following nine problem descriptions, literally copied from different case histories, were combined to one description, i.e., "fantasizes much, tells fantasy stories": (a) much fantasizing, (b) tells fantasy stories, (c) cannot separate fantasy and reality, (d) mixes fantasy and reality, (e) implausible stories, (f) tells many fantasy stories, (g) has a lot of fantasy. These aggregated descriptions represent problem behaviors presumably belonging to the domain covered by the CBCL but not listed in it. These aggregated descriptions will be referred to as non-assigned items. We checked how often the CBCL items and the *nonassigned* items appeared in the 93 reports (i.e., the prevalence rate of the two types of items).

Reliability

To study the reliability of the procedure just described, we randomly divided the 93 reports into three groups of 31. Three raters, recent graduates in education (Rater B), clinical psychology (Rater C), and research psychology (Rater D), each independently analyzed a set of 31 reports. Guided by a written instruction illuminating the aforementioned procedure, they also extracted the problem behaviors from the reports and compared them with the CBCL items. Their results were compared with that of the author (Rater A) who had analyzed all 93 reports.

The percentage of agreement between Rater A and the other three raters was determined for the extraction of the descriptions from the reports, and for the comparison of these problem descriptions with the CBCL. 85%, 82% and 89% of the behavior problems extracted by Rater A were also extracted by Rater B, C and D, respectively.

The average percentage of agreement for Rater A with the other three raters with respect to the assignment of reported problems to CBCL was .77. The agreement between A and B was .83, between A and C it was .75 and between A and D it was .71. We only checked whether two observers agreed on assigning an extracted problem behavior to the CBCL, not whether they assigned it to the same CBCL item. Our main purpose was to check whether an extracted problem behavior could or could not be assigned to the CBCL. If two raters assigned an extracted problem behavior to different CBCL items, at least they agreed that it belongs to the domain covered by the CBCL.

3.2.2 Results

From the 93 reports, 811 problem behaviors were extracted by Rater A. 687 of these were also extracted by Rater B, C, and D together. The rate of assignment of these 687 behaviors to the CBCL is shown in Table 3.2, for each rater pair separately as well for Rater A with Raters B,C and D combined.

Table 3.2
Comparisons of the assignment of the extracted problem behaviors to the CBCL between Rater A and the other three raters.

		Rater A		
		assigned	not-assigned	
Rater B	assigned	134	35	169
	not assigned	7	71	78
		141	106	247
Rater C	assigned	118	44	162
	not assigned	13	56	69
		131	100	231
Rater D	assigned	99	41	140
	not assigned	19	50	69
		118	91	209
Raters B,C, and D together	assigned	351	120	471
	not assigned	39	177	216
		390	297	687

The table shows that 71 of the 247 reported behaviors judged by both raters A and B (i.e., 29%) were not assigned by either of them to any CBCL item. The corresponding figures for Raters A and C are 56 (24%), and for Raters A and D 50(24%). In all, 177(26%) of the 687 extracted problem descriptions were not assigned to the CBCL by either of the raters who judged them.

Of these 177 problem descriptions, 30 were mainly related to a school situation or were considered to concern a specific learning problem. On the basis of the remaining 147 descriptions, 47 nonassigned items were formulated, each of which characterizes one or more of the nonassigned problem behaviors. The description of these items is given at the end of Table 3.3, which also shows (in the first column) the prevalence of each item.

Table 3.3
Prevalence rate of the CBCL items and the 47 nonassigned items in different problem behavior samples.*

		NVO reports	OQI	Expanded CBCL
CBCL items:				
1.	Acts too young for his/her age	10	0	17
2.	Allergy	0	0	3
3.	Argues a lot	20	6	13
4.	Asthma	0	0	2
5.	Behaves like opposite sex	1	0	1
6.	Bowel movements outside toilet	4	0	4
7.	Bragging, boasting	3	1	15
8.	Can't concentrate, can't pay attention for long	43	3	22
9.	Can't get his/her mind off certain thoughts:obsessions	1	1	13
10.	Can't sit still, restless, or hyperactive	27	0	12
11.	Clings to adults or too dependent	5	0	11
12.	Complains of loneliness	0	0	5
13.	Confused or seems to be in a fog	3	0	4
14.	Cries a lot	8	2	1
15.	Cruel to animals	0	1	1
16.	Cruelty, bulling, or meanness to others	0	0	2
17.	Daydreams or get lost in her/her thoughts	10	0	7
18.	Deliberately harms self or attempts suicide	1	0	0
19.	Demands a lot of attention	16	1	16
20.	Destroys his/her own things	1	0	1
21.	Destroys things belonging to his/her family or other children	2	0	1
22.	Disobedient at home	13	10	8
23.	Disobedient at school	1	0	4
24.	Doesn't eat well	6	4	9
25.	Doesn't get along with other children	10	1	4
26.	Doesn't seem to feel guilty after misbehaving	0	0	5
27.	Easily jealous	2	0	7
28.	Eats or drinks things that are not food	0	0	0
29.	Fears certain animals, situations or places other than school	4	0	8
30.	Fears going to school	0	0	0
31.	Fears he/she might think or do something bad	1	0	3
32.	Feels he/she has to be perfect	2	0	13
33.	Feels or complains that no one loves him/her	0	0	1
34.	Feels others are out to get him/her	4	0	2

** The first column represents the sample from the 93 NVO reports. The second column represents the sample from the open-question inventory (OIQ) at 28 parents of physically handicapped children. The third column represents the inventory from the expanded CBCL at the same 28 parents.*

		NVO reports	OQI	Expanded CBCL
35.	Feels worthless of inferior	0	0	5
36.	Gets hurt a lot,accident-prone	3	0	5
37.	Gets in many fights	8	1	2
38.	Gets teased a lot	8	0	6
39.	Hangs around kinds who get in trouble	1	0	0
40.	Hears things that are not there	1	0	0
41.	Impulsive or acts without thinking	9	0	13
42.	Likes to be alone	2	0	13
43.	Lying and cheating	6	0	3
44.	Bites fingernails	2	0	1
45.	Nervous, high-strung, or tense	13	0	6
46.	Nervous movements or twitching	2	0	4
47.	Nightmares	0	0	3
48.	Not liked by other children	1	0	3
49.	Constipated, doesn't move bowels	0	0	3
50.	Too fearful or anxious	9	0	4
51.	Feels dizzy	0	0	1
52.	Feels too guilty	0	0	1
53.	Overeating	0	0	3
54.	Overtired	1	0	3
55.	Overweight	0	0	6
56.	Physical problems without known medical cause;			17
	a. aches or pains	1	1	
	b. headaches	3		
	c. nausea, feels sick		0	
	d. problems with eyes	3		
	e. rashes of other skin problems	0		
	f. stomachaches or cramps	3		
	g. vomiting, throwing up	0		
	e. other	0		
57.	Physically attacks people	4	1	0
58.	Picks nose, skin, or other parts of body	0	0	0
59.	Plays with own sex parts in public	0	0	0
60.	Plays with own sex parts too much	0	0	2
61.	Poor school work	20	0	8
62.	Poorly coordinated or clumsy	4	1	11
63.	Prefers playing with older children	0	0	6
64.	Prefers playing with younger children	4	0	14
65.	Refuses to talk	1	0	3
66.	Repeats certain acts over and over, compulsions	0	0	3
67.	Runs away from home	0	0	1
68.	Screams a lot	2	3	3
69.	Secretive, keeps thing to self	14	0	10
70.	Sees things that are not there	0	0	0
71.	Self-conscious or easily embarrassed	0	0	8
72.	Sets fire	1	0	0
73.	Sexual problems	0	0	0
74.	Showing off or clowning	3	1	13
75.	Shy or timid	6	12	7
76.	Sleeps less than most children	0	0	5
77.	Sleeps more than most children during days and/or night	0	0	3
78.	Smears or plays with bowel movements	1	0	0
79.	Speech problems	9	1	10
80.	Stares blankly	0	0	3
81.	Steals at home	5	0	1
82.	Steals outside the home	3	0	1
83.	Stores up things he/she doesn't need	0	0	2
84.	Strange behavior	9	0	2
85.	Strange ideas	0	0	1
86.	Stubborn, sullen, or irritable	4	1	10
87.	Sudden changes in mood or feelings	1	1	6
88.	Sulks a lot	2	7	9
89.	Suspicious	0	0	1
90.	Swears or uses obscene language	1	2	5
91.	Talks about killing self	1	0	0

		NVO reports	OQI	Expanded CBCL
92.	Talks or walks in sleep	0	0	3
93.	Talks too much	0	0	14
94.	Teases a lot	3	1	7
95.	Temper tantrums or hot temper	14	3	10
96.	Thinks about sex too much	0	0	0
97.	Threatens people	0	0	0
98.	Thumb-sucking	1	0	4
99.	Too concerned with neatness or cleanliness	0	0	6
100.	Trouble sleeping	9	2	3
101.	Truancy, skips school	2	0	0
102.	Underactive, slow-moving, or lacks energy	8	0	8
103.	Unhappy, sad of depressed	7	0	2
104.	Unusually loud	3	0	6
105.	Uses alcohol or drugs	0	0	0
106.	Vandalism	0	0	0
107.	Wets self during the day	1	2	2
108.	Wets the bed	5	2	4
109.	Whining	0	3	2
110.	Wishes to be of opposite sex	0	0	0
111.	Withdrawn, doesn't get involved with others	16	0	2
112.	Worrying	4	0	3
Added Items				
1.	Fantasizes a lot, tells fantasy stories	4	0	8
2.	Messy work	3	1	7
3.	Fails to finish things he/she starts	3	1	16
4.	Defiant, impudent answers	4	1	9
5.	Strongly opinionated, talks back often	1	0	11
6.	Forgetful, can't remember things	10	2	8
7.	Is or plays alone often	2	0	22
8.	Needs constant supervision during play or work time	1	1	6
9.	Defiant, refuses to do what he or she is told	1	7	4
10.	Inconsiderate to others, insensitive	1	0	7
11.	Hates school, doesn't like to go to school	1	0	2
12.	Afraid to make mistakes, fear of failure	4	0	17
13.	Behaves irresponsibly	1	0	3
14.	Shows few emotions	1	0	7
15.	Sneaky	1	2	3
16.	Unusual playing behavior	2	2	3
17.	Gets easily angry if she or he can't have her or his own way	6	5	20
18.	Impatient	1	1	15
19.	Gives in easily, not assertive	3	3	10
20.	Is not affected by rules or norms	3	1	5
21.	Has difficulty adapting	1	0	8
22.	Imitates others	1	2	8
23.	Escapes from certain situations and/or problems	1	1	14
24.	Bossy	8	2	9
25.	Has difficulties with motor movements	4	3	23
26.	Doesn't do homework	2	1	2
27.	Irritable	1	0	7
28.	Lacks self-confidence	4	0	15
29.	Avoids physical contact	1	0	3
30.	Needs help with eating, washing, and/or dressing	2	2	17
31.	Panics easily	2	2	12
32.	Materialistically oriented	1	1	3
33.	Continues trying to have things his or her own way	1	9	17
34.	Rejects help from mother	1	2	3
35.	Doesn't participate in competitive games	1	0	9
36.	Doesn't seem to be afraid, shows no fear	1	0	15
37.	Can't handle demands	2	1	12
38.	Sensitive, easily hurt	2	0	22
39.	Hesitates often, can't make up his or her mind or make choices	1	0	12
40.	Naughty	2	0	20
41.	Minds everyone's business	1	0	13

		NVO reports	OQI	Expanded CBCL
42.	Homesick	2	0	7
43.	Talks a lot about death	1	0	9
44.	Can't stand losing	1	1	19
45.	Strongly wants contact with others	1	0	24
46.	Behaves aggressively toward certain persons in his or her environment	1	0	3
47.	Can't stand being alone	1	3	9

These results show that in psychodiagnostic practice many problem behaviors are reported that are not represented in the item sample of the CBCL.

3.3 Study 2: Content validity of an expanded CBCL[1]

This study was designed to establish whether the problem behaviors that appear in a specific clinical group, namely children with motor dysfunctions, can be determined with the CBCL and the 47 nonassigned items. Furthermore, the added value of the CBCL method above an open-question method during the diagnostic phase of problem identification was analyzed.

3.3.1 Method

In this study 28 parents of children with motor dysfunctions aged 6-14 took part. On the basis of the open question "which behaviors of your child do you think are problematic", parents gave written reports of problem behaviors. About 2 months later the same parents filled in the CBCL, expanded with the aforementioned 47 items. The problem behaviors that emerged from this method were compared with the problem behavior descriptions that emerged from the open-question method (for more details we refer to Ten Cate & Hoge, 1989).

3.3.2 Results

140 behaviors were reported by the parents with the open-question method. From these behaviors 20 could not be scored in the expanded CBCL. After similar behaviors were combined 13 problem behaviors remained. In Table 3.4 the open-question method and the expanded CBCL method are compared in terms of the number of problems found.

Results show that more than six times as many problem behaviors were inventoried with the expanded CBCL (975) as with the open-question method (140). Of the total amount of inventoried problem behaviors (975+140), regardless of the method, the expanded CBCL covered over 87%. The CBCL without the 47 supplementary items covered 48%. Column 3 of Table 3.3 shows

which CBCL items and nonassigned items were filled in by the parents. Twelve items of the CBCL were filled in by more than thirteen parents. Fifteen of the 47 supplementary items were also filled in by more than thirteen parents.

Table 3.4
Comparison of the number of inventoried problem behaviors using the open-question method, the CBCL and the expanded CBCL

	Number	
Behaviors inventoried with the open-question method	140	
Represented in expanded CBCL		120
Not represented in expanded CBCL		20
Behaviors inventoried with the expanded CBCL	975	
Filled in on the standard CBCL items		507
Filled in on the 47 supplementary items		468
Total number of inventoried behaviors	1115	

3.4 Discussion

This two studies demonstrate that the items of the CBCL do not cover the broad range of problem behaviors that are mentioned in diagnostic reports. Of course, no checklist can ever represent every exact symptom which might possibly occur to any child. The dysfunctionality of behavior depends on cultural and sociological conditions and varies over time. But we may question whether the CBCL is representative enough to be used as a data base in a computerized diagnostic knowledge system.

It is interesting to speculate about possible explanations for these findings. One can argue that the problem behaviors collected in this study are not relevant with respect to the content validity of the CBCL. Is it justified to use diagnostic reports as an information source to determine the content validity of an instrument that parents fill in themselves? To what extent do the reports reflect what the diagnosticians consider to be a problem, and not what the problem is according to the parents? But the fact that the reports came from professional registered diagnosticians give us confidence in the accuracy of the reproduction of the problems parents had reported to them. The registration committee requires a concrete description of the problems of the client in the diagnostic report. Note also that the items of the CBCL were developed through an extensive study of child psychiatric case histories (Achenbach, 1966).

In this study, problem behaviors were not collected in natural settings of mental health services. Instead, we presumed that the file of case histories of a Dutch organization of diagnosticians would represent the diversity of the diagnostic practice. It is difficult to check the representativeness of this file since

there is no appropriate overview of mental health services for children with behavior problems. However, even if the file does not reflect the diversity of problems existing in the clinical field, it does not affect the conclusion that the CBCL, at least for Dutch clinically referred children aged 6-11, does not represent all the relevant problem behaviors.

All problem behaviors collected in this study were reasons for clinical referral. Some of the nonassigned or supplemental items may look unimportant or infrequent to many a diagnostician. But that is also true for some CBCL items. Achenbach himself remarks that "many of the problem behaviors are not intrinsically pathological, but are exhibited to some extent by most children and youth" (1985, p.17). The behavior becomes problematic by virtue of the frequency with which it is exhibited by the child. In column 1 of Table 3.3 one can see that 43 CBCL items did not appear as problematic behavior in the report sample. 19 CBCL items appeared once in the report sample.

It is difficult to decide about the clinical relevance of problem behavior. In the reports written by the parents fewer problems were always mentioned than when, the CBCL was filled in. One can infer from this fact that probably in the reports only the most striking and therefore relevant problem behaviors were mentioned. During the phase of identification of problem behaviors it is important not to dismiss information too quickly as being irrelevant. Column 3 of Table 3.3 shows that the added items were frequently filled in by the parents of the children with motor dysfunctions. Further empirical research should provide an answer to the question of the discriminative validity of the added items.

While preparing this manuscript we were informed by T.M. Achenbach and F.C. Verhulst (personal communication, June 1989) that Achenbach, Conners and Quay have constructed an expanded behavior checklist (ACQ behavior checklist for ages 4-16) which contains one hundred items more than the CBCL. We can only agree that our empirical findings also strongly support an expansion of the item base of the CBCL and the need for empirical research comparing this list with the CBCL. This does not imply that we question the factor structure of the CBCL, as revealed in its empirical syndromes. But it raises the question whether the added items or symptoms are part of the syndromes that have already been empirically derived. This question can only be answered by research in which old and new items are submitted to factor analysis.[2]

For now it can be concluded that if the CBCL is intended to be used as a standardized assessment instrument in a knowledge system, according to our findings a number of items should be added.

3.5 Implications for further research

Results from this study show that the clinician regularly is confronted with individual problem behavior that is not mentioned in an empirical classification

system such as the Achenbach system. One can advocate empirical research to determine whether these problem behaviors or symptoms are part of the syndromes that have already been empirical derived. But this approach cannot solve the more important functional problem in the use of classification systems, namely how to deal with individual problem behavior that doesn't fit into the classes of the system. How does or can a diagnostician decide whether the symptom or problem behavior is part of a class or a syndrome? To obtain practical guidelines for solving this classification problem, we focused the rest stage of our research on the classification processes of experienced diagnosticians. How do diagnosticians classify behaviors, and in particular, the special cases not found in an actuarial system? A study on psychodiagnostic classification will be reported in the next chapter.

1. We especially want to thank Dr. E. van Aarle for providing the idea of examining the expanded checklist in a practical diagnostic situation and for his guidance during the execution of this study.

2. Some of the questions raised here have in the meantime been more or less answered, because Achenbach has reported that empirical research with the ACQ failed to identify any syndromes that were not identified by the CBCL. Furthermore, the ACQ discriminated less well than the CBCL between referred and non-referred samples (Achenbach, 1989). To our best knowledge the co-occurrence of the added items or symptoms with the already empirically derived symptoms has not been reported in the literature.

4 Diagnostic classification by experts and novices

4.1 Introduction

Classification of dysfunctional behavior is an important task in clinical decision making (Achenbach, 1985; De Bruyn, 1985; Rispens, 1986; Rutter, Tuma, & Lann, 1988). Whether the diagnostician intends this or not, it focuses or guides the generation of hypotheses concerning the conditions that elicit or sustain the identified problem. It narrows down the field in terms of causes, treatment, and prognosis and, as noted by Rutter (1975), it provides a short-hand language for communication with other professionals.

Several classification systems have been proposed to systemize and support this diagnostic task, for example DSM-III-R (1987) and the CBCL system (Achenbach & Edelbrock, 1983). Whatever classification system the diagnostician is using, he or she will inevitably be unable to classify all observed problematic behaviors with the system at hand. Krol and De Bruyn (1990a;1990b), for example, screened a nationwide sample of diagnostic files for characterizations of behavior problems in children. Forty seven problems, many of which were frequently mentioned in the files, were not included in the standard list of Achenbach's classification system (cf. the previous chapter).

One might argue that it is not feasible in principle to design classification systems that cover all manifestations of problem behavior. The definition, the dysfunctionality, as well as the manifestation of problem behavior seem to depend both on conditions in nature and on historical, cultural and sociological conditions. Therefore, diagnosticians will inevitably observe instances of problem behavior that are missing in the database of the classification system at hand. This does not imply that it is impossible to formulate decision rules that may help the diagnostician to solve such a classification problem. But where do such rules come from?

When diagnosticians are confronted with instances of dysfunctional behavior not included in a classification system, they have to solve the problem on their own. If we assume that experienced or expert diagnosticians have developed internal classification rules, we can try to extract these rules from them, formalize these rules, and add these to the existing classification system. Such an approach was followed as early as in 1968 by Kleinmuntz, who was able to extract the decision rules of an expert classifying MMPI profiles. According to Kolodner (1984), the ability of experts to deal with exceptional or novel cases develops as a result of noticing failures and successes, as well as differences and similarities between "cases". Her model of the development of expertise proposes two steps in the evolution from novice to expert. "First, knowledge is built up incrementally on the basis of experience. Facts, once unrelated, get integrated through occurrence in the same episodes. Second, reasoning processes are refined, and usefulness and rigidity of rules is learned" (1984, p.96). Experience reorganizes the structures of both the reasoning process and the domain knowledge

According to Kolodner's model, experience with the "art" of classification, is continously refining the reasoning process of diagnosticians. As a consequence, we expect to find differences between experts and novices in their reasoning processes during classification. To get a usable description of this reasoning process, we use the method of analyzing thinking-aloud protocols. This method seems to be particularly suited for detecting implicit cognitive operations (Ericsson & Simon, 1984). Our study focuses on the description of the cognitive operations involved in classification and investigates whether experts and novices differ in their use of cognitive operations during the classification of behavioral problems.

The model of Kolodner also implies that experience, or the observation of co-occurrences of symptoms in "cases", reorganizes the relationships between objects (i.e., symptoms and categories) in the domain knowledge. With respect to the field of clinical diagnostics, this means that experts should differ from novices in both the range of syndromes in which they organize symptoms, and in the ways in which symptoms within a syndrome are related to each other. In turn, differences in the range and internal organization of syndromes will lead to differences in the extent to which a dysfunctional behavior unit reflects the character of a syndrome, that is, the extent to which a particular symptom is typical for a syndrome. Thus, the second question we address in this study is whether experts and novices differ with respect to the outcome of the classification, in particular the typicality ratings of symptoms.

In studies on decision making, task effects have often been demonstrated (Hamm, 1988). Therefore, we also investigate whether the reasoning processes of experts are influenced by the nature of the task: we compare the reasoning processes on two classification tasks, classifying new symptoms into a given classification system versus intuitive clustering of symptoms.

To summarize, this chapter addresses the following questions: a) Do experts and novices differ in their use of cognitive operations during the classification of behavioral problems?, b) Do experts and novices differ with respect to the outcome of the classification, in particular the rating of the typicality of symptoms? and c) Does the nature of the classification task affect the use of cognitive operations?

	Task			
	A	B	C	N
Expert group 1	x	x		(n=12)
Novice group	x	x		(n=12)
Expert group 2			x	(n=10)

Figure 4.1: The design of the study
A: Classifying new symptoms in terms of CBCL syndromes; B: Rating new symptoms with respect to CBCL syndromes; C: Intuitive clustering of the original CBCL-symptoms.

4.2 Method

Design and Subjects

The study involved three groups of subjects and three tasks. The assignment of tasks to groups is shown in Figure 4.1.

The first group of experts consisted of 12 diagnosticians, selected with the help of a list of all of the institutions in the Netherlands involved in mental health services. They had been engaged in clinical practice for several years and were familiar with the diagnosis of behavior problems in children 6 to 11 years old. Six different work settings were chosen which are representative of the entire mental health field. From each setting two diagnosticians were selected, one clinical psychologist and one special educator, one male and the other female (Table 4.1).

Table 4.1
Composition of the two groups of experts

Expert Group 1		Experience in years
Setting		
Mental Health Services (RIAGG)	1 Male Psychologist	19
	2 Female Special Educator	12
Hospital for children or Medical Child Institution	3 Female Psychologist	17
	4 Male Special Educator	8
Special education treatment institution	5 Male Psychologist	25
	6 Female Special Educator	27
Clinic for child psychiatry	7 Male Psychologist	16
	8 Female Special Educator	7
School Consultation Center	9 Female Psychologist	12
	10 Male Special Educator	16
Remaining (Day care center)	11 Female Special Educator	10
	12 Male Psychologist	2
Expert Group 2		
Setting		
Hospital for children or Medical Child Institution	1 Female Psychologist	9
	2 Male Special Educator	5
Special education treatment institution	3 Male Psychologist	9
Clinic for child psychiatry	4 Female Psychologist	8
	5 Male Psychologist	8
	6 Male Special Educator	10
	7 Male Special Educator	17
	8 Male Psychologist	3
	9 Male Special Educator	11
	10 Female Special Educator	4

A second group of subjects consisted of novices: 12 female students at the University of Nijmegen, 6 majoring in Psychology and 6 in Special Education. Starting from the beginning of an alphabetical list (in the case of Special

Education) or from the end (Psychology) students were invited by telephone to participate, until the required number of subjects was reached. There was also a second group of experts which was similar to the first one but not quite as representative with respect to work setting and gender (see Table 4.1). None of the experts in the first group was using the Child Behavior Checklist (CBCL) in his or her work, although most of them did know of its existence. As to the novice group, the CBCL had not yet been studied in their courses. In the second expert group, only one diagnostician was using the CBCL system in his practice (the other 9 subjects were not familiar with the CBCL).

Classification tasks and procedure

The three tasks used in this study were all concerned with classifying symptoms in one way or another. In the construction of the tasks we used the empirical classification system as developed by Achenbach because it is the best validated and reliable system in the field of child psychopathology (Achenbach, Verhulst, Baron, & Althaus, 1987; cf. chaps. 2 and 3 of this thesis). This classification system is based on ratings obtained with the Child Behavior Checklist (CBCL), a parents' report on the problem behavior of their child.

Taks A: Classifying new symptoms in terms of CBCL syndromes.

In Task A, 'new' symptoms had to be classified by means of the syndrome system provided by the CBCL. Fourty-seven problem behaviors that were observed in clinical practice but that are not included in the CBCL system served as new symptoms (cf. Krol & De Bruyn, 1990, and the description of Study 1 in chap. 3 of this thesis. The 47 items are listed at the end of Table 3.3). The problem behaviors were printed on small cards, the CBCL syndromes on larger cards. Each of the syndrome cards listed the CBCL items defining that syndrome, printed in a random order.

There were seven cards for the syndromes of boys and seven for those of girls. The labels of the syndromes were not shown on the cards. One miscellaneous card contained CBCL items that do not occur together in any single syndrome. Six subjects of each group received the girls' syndromes plus the miscellaneous card and six the boys' syndromes plus the miscellaneous card.

Subjects were instructed to consider whether, in their experience, the problem behaviors on the small cards belonged to any of the syndromes. They were instructed to think aloud during the task. The role of the interviewer was nondirective, although she prompted the subjects to verbalize their thoughts. As soon as the task was completed, the subjects were asked to answer a few questions concerning their education, their place of work, and their number of years of experience.

The thinking-aloud protocols of Task A provided information with respect to the cognitive processes involved in symptom classification. The outcomes of Task A provided information with respect to the embedding of problem behaviors in a

well-defined empirical classification system.

Task B: Rating new symptoms with respect to CBCL syndromes.
In Task B the syndrome cards and the instructions were changed, partly for reasons not relevant to the present study. The items defining the syndromes were printed on the cards in descending order of factor loading; the factor loadings themselves were not printed on the cards (see Table 4.2 for an example card and Appendix A for a listing of all syndromes in Dutch).

Table 4.2
Example of a syndrome card where the CBCL items defining the syndrome are listed in descending order of their factor loading

Syndrome card 4		Factor loading*
1	Acts to young for her age	65
62	Poorly coordinated or clumsy	65
61	Poor school work	58
8	Can't concentrate, can't pay attention for long	55
41	Impulsive or acts without thinking	48
64	Prefers playing with younger children	41
13	Confused or seems to be in a fog	40
38	Gets teased a lot	38
10	Can't sit still, restless, or hyperactive	38
79	Speech problems	35
23	Disobedient at school	33
102	Underactive, slow moving, lacks energy	33
17	Day-dreams or gets lost in her thoughts	31

* *The factor loadings were not written on the cards*

To control for memory effects subjects were given the syndromes from the gender group that were not used in Task A. The subjects received the syndrome cards one by one, in random order. They had to indicate a "typicality rating" for each problem behavior in relation to each syndrome on a scale from 0 to 9, with '9' meaning that a problem behavior is very much a part of the syndrome and '0' that it does not belong to the syndrome at all. Subjects were instructed to refrain from thinking aloud. The outcomes of Task B, like those of Task A, also provided information with respect to the embedding of problem behaviors in a well-defined empirical classification system.

Task C: Intuitive clustering of the original CBCL symptoms.
In Task C expert diagnosticians were asked to cluster the original items of the CBCL from their own points of view. Subjects were first given a set of 112 small cards containing the problem behaviors and child characteristics as used by the CBCL. From this set they could select the behaviors and child characteristics with which they were familiar. Problem behaviors and characteristics which they met in their practice but which were not in the CBCL set, could be written down on new cards. Next, subjects were asked to sort the collected behaviors and characteristics into clusters containing items that in their own clinical experience frequently co-occurred. They were allowed to form as many clusters as they

wanted and to put the same item in more than one cluster. They were instructed to think aloud during this task. In this task the question is whether the same cognitive processes are involved in sorting the original CBCL items into clusters and in classifying new symptoms into the CBCL system. The question of whether the outcome of the sorting process is in agreement with the empirically established system of clusters is dealt with in chapter 5.

Table 4.3
Definitions of cognitive operations used in a psychodiagnostic classification task.

1. Asking for or giving information
Questions about or remarks on the problem behavior, the syndrome or items of a syndrome, e.g.
'I don't consider "naughty" that to be a problem'
'I don't understand what is meant, the behavior is not concrete'
'The description of the behavior is too vague'

2. Associating
Looking for or mentioning associated behaviors and/or characteristics. This is done by using phrases like 'the child will also have behavior x and characteristic y', or by talking about an image of a child, or by giving an example of a child. These 'associations' are formulated as concrete behaviors or characteristics, e.g.
'What kind of child panics easily? well I think a child who is afraid of losing control over the situation'
'This child goes his own way' (= Syndrome 1)
'This child has a very good memory' (= Syndrome 2)

3. Abstracting or labeling
Assigning a label, an abstraction or an interpretation to a problem behavior, syndrome or item of a syndrome. This is done by referring to a problem group, e.g., children with attachment problems or contact disorder, or by mentioning a personality type, e.g. neurotic. Unlike 'associations' these labels or abstractions are not concrete behaviors. This cognitive operation is concerned with the grouping or ordering of behaviors, characteristics or syndromes under a broader denominator, e.g.
'You often see this (i.e., the problem behavior) in children found in the domain of psychiatry'
'These (the problem behaviors) all have to do with social contact'
'This is the neurotic syndrome' (Syndrome 3)
'I have put syndromes 1 and 2 together, they are all extravert behaviors and those are the introvert syndromes 7, 5, and 3'

4. Explaining
Explaining a behavior, a syndrome or an item of a syndrome by mentioning reasons or causes for the problematic behavior or syndrome. This operation is concerned with the explanation of problems and syndromes, e.g.
'I am looking for the reasons of this behavior'
'These are behavior problems caused by a lack of structure'
'You will often see this conglomeration of behaviors (Syndrome 2) when a fundamental disturbance in the relation between mother and child is involved'

5-7. Matching
Matching: comparing the different concepts with each other. Phrases like 'this looks like', 'this is the same as', 'this belongs to' and 'you see this behavior in' are used.
The following distinction is made concerning the result of the matching:

5. Neutral Matching
Comparing of concepts without a definite conclusion. It is not clear if the concepts match or do not match, e.g.
'I am wondering between syndrome 6 or 5 is applicable'

6. Identifying (positive match)
Identifying, comparing of concepts resulting in a positive match
'That (i.e., problem behavior) belongs to syndromes 1 and 2'

7. Differentiating (negative match)
Differentiating, comparing of concepts resulting in a negative match
'I don't think that he panics easily, number one' (i.e., Syndrome 1)

Note. In the coding system, the names of the cognitive operations were omitted to encourage the coders to look at the content, i.e. the definitions of the operations.

Tasks A and B were presented in two seperate sessions, three weeks apart. Task A preceded Task B, so as to obtain classification results that were unaffected by the instructions and materials of Task B. For the experts, a two-hour time-limit was in effect for Task A; a few subjects did not complete the task on time, but some of them went on to complete it anyway. Novices were likewise not allowed more than two hours for Task A, and three of them did not complete it. All subjects completed Task B and C on time. Sessions with the diagnosticians were conducted at their own offices, whereas sessions with the novices were conducted at the University. The first author administered Task A and B, whereas two instructors administered Task C at the offices of the diagnosticians.

Coding the thinking-aloud protocols

As yet, no category system has been accepted to describe the cognitive operations involved in the part of the diagnostic process that involves classifying problem behaviors. We used the generic model of classification called heuristic classification (Breuker, 1986; Clancey, 1985) during the development of our category system. Seven categories were distinguished (see Table 4.3).

To identify the cognitive processes involved in classification, we analyzed the verbal protocols from Tasks A and C using the guidelines of Ericsson and Simon (1984). The protocols were transcribed verbatim. The analysis of the protocols involved two activities. First, we segmented the text into units. A segment or unit was defined as a piece of text containing enough information to associate it with a code. Second, we assigned codes corresponding to cognitive operations (see Table 4.3). When a segment was insufficiently self-contained, for example because it included anaphoric references like 'this' or 'that', at most two immediately preceding or following segments were consulted in order to obtain the information required for coding it (for details, see Krol, 1990).

The reliability of the system was checked by having two persons code four randomly selected pages of each protocol. Then, kappa (Cohen, 1960) was used as an index of agreement. Agreement with respect to segmenting a protocol section into coding units ranged from .61 to .95, with a mean of .86 over all 22 protocols. The kappa values for agreement in coding cognitive operations ranged from .44 to .97, with a mean of .65; sixteen values were "moderate" (between .41 and .60) , thirteen were "substantial" (between .61 and .80), and five were "almost perfect" (between .81 and 1.00)(Landis & Koch, 1977; Popping ,1983).

Data analysis

The first expert group and the novice group were compared with respect to the cognitive operations they used in classifying symptoms in Task A, first by means of a univariate analysis of variance with the total frequency of operations as the dependent variable, and second, by means of a multivariate analysis of variance involving the relative frequencies of the different operations. A similar comparison was made between the first and the second expert group to reveal

potential task effects.

To compute the agreement between the expert and the novice group with respect to the outcome of Task A, (the classification of 47 problem behaviors into CBCL syndromes), we first determined for each problem behavior a mean classification score for each group by counting the number of times an expert or novice classified it in terms of a specific syndrome and divided this sum by the number of subjects in each group. This score indicated which CBCL syndrome the behavior belonged to according to the expert group (n=10) or according to the novice group (n=9). The ns are lower than the number of subjects per group (12) because the scores of the subjects who did not complete Task A are not included.

Most subjects in both groups classified a given problem behavior in terms of more than one syndrome. Then for each syndrome we correlated the 47 mean classification scores of the expert group with the scores of the novice group (using the Pearson product-moment correlation).

Table 4.4
Relative frequencies(in %) of the cognitive operations for each expert and novice*

	Expert group 1, Task A											
Cognitive operation	1	2	3	4	5	6	7	8	9	10	11	12
1. Asking	10.8	7.3	6.3	1.9	13.6	8.1	3.9	12.4	10.0	7.8	14.4	13.3
2. Associating	23.3	22.2	17.5	10.5	19.2	18.6	14.0	12.7	10.8	6.0	5.6	12.3
3. Labeling	13.3	25.5	12.4	13.1	7.7	12.0	16.4	15.4	18.9	11.8	26.6	10.0
4. Explaining	4.1	5.8	4.7	2.1	10.7	3.2	4.1	2.7	0.9	0.6	10.9	2.4
5. Neutral Matching	17.7	23.3	13.4	40.3	30.8	33.8	28.9	39.9	39.3	40.6	21.0	31.2
6. Identifying	24.1	13.9	30.0	22.6	14.2	20.7	17.0	9.3	18.5	20.5	19.6	20.3
7. Differentiating	6.8	2.0	15.7	9.5	3.8	3.5	15.9	7.5	1.6	12.7	1.6	9.7
	Novice group, Task A											
Cognitive operation	1	2	3	4	5	6	7	8	9	10	11	12
1. Asking	3.6	3.4	4.5	3.7	2.9	7.1	1.5	9.6	2.3	5.2	2.2	4.8
2. Associating	15.7	2.8	9.8	11.3	12.4	17.4	11.0	12.8	8.7	9.0	5.7	2.5
3. Labeling	6.9	3.4	2.6	7.5	7.3	9.4	1.0	4.9	4.4	6.8	10.7	4.0
4. Explaining	0	0	0	0.3	0.9	0	0.1	0	1.3	1.0	0	0
5. Neutral matching	27.3	61.4	40.5	40.1	41.8	44.2	29.2	44.5	42.3	36.2	52.1	38.5
6. Identifying	20.4	20.2	36.1	25.2	22.0	14.3	24.4	21.6	14.1	20.9	8.0	19.1
7. Differentiating	26.0	8.9	6.5	12.0	12.7	7.5	32.7	6.7	26.9	20.9	21.2	31.2
	Expert group 2, Task C											
Cognitive operation	1	2	3	4	5	6	7	8	9	10		
1. Asking	12.9	5.1	3.5	9.6	6.0	13.1	12.9	7.9	16.1	13.1		
2. Associating	12.9	2.5	13.9	22.3	10.0	23.0	12.5	7.1	5.2	26.6		
3. Labeling	39.4	84.6	74.1	39.9	62	37.6	59.6	65.9	61.8	43.2		
4. Explaining	7.6	1.3	6.1	6.1	6.0	6.9	3.5	11.7	14.6	3.5		
5. Neutral matching	0.0	0.4	0.3	1.9	0	0.6	1.0	0.8	0.4	0.9		
6. Identifying	22.7	4.3	0.9	15.3	16.0	12.8	6.8	2.4	1.1	6.6		
7. Differentiating	1.5	1.7	1.2	5.0	0	6.0	3.9	4.3	0.7	6.1		

**The percentages refer to the number of times a segment was assigned to each category divided by the total number of segments in the protocol (i.e., the total number of operations).*

Agreement between experts and novices regarding the typicality ratings of the 47 new problem behaviors with respect to the CBCL syndromes was computed as follows: For each problem behavior and each of the boys' syndromes, six ratings were available in each group; and similarly for the girls' syndromes. For each syndrome, the 47 mean typicality ratings of the experts and the 47 mean typicality ratings of the novices were correlated using the Pearson product-moment correlation, which provided an overall index of agreement in typicality. Furthermore, we applied a t-test to the two mean typicality scores for each problem behavior, separately.

Finally, for each syndrome we correlated the mean classification score of each problem behavior (from Task A) with the mean typicality score of each problem behavior (from Task B). In this way we obtained an index of consistency of the judgements over the two tasks.

The actual embedding of the problem behaviors into the empirical classification system is summarized.

4.3 Results

Cognitive operations of experts and novices during the classification of new symptoms in terms of CBCL syndromes.

The relative frequency of each cognitive operation for each subject is given in Table 4.4.

Table 4.5 shows the mean relative frequencies of the seven cognitive operations as a function of experience (experts versus novices) and as a function of framing (Task A versus Task C).

Table 4.5
Mean relative frequency of cognitive operations as a function of expertise and task.

	cognitive operations in Task A						
group	1	2	3	4	5	6	7
expert group 1	9.2	14.4	15.2	4.4	30.0	19.2	7.5
novice group	4.2	9.9	5.7	.3	41.5	20.5	17.7
univariate p<.05	*		*	*	*		*
	cognitive operations of experts						
task	1	2	3	4	5	6	7
task A (expert group 1)	9.2	14.4	15.2	4.4	30.0	19.2	7.5
task C (expert group 2)	10.2	13.6	56.8	6.7	.6	8.8	3.0
univariate p<.05			*		*	*	

The univariate analysis of variance with the total frequency of operations in Task A as dependent variable showed no significant difference between the expert group and the novice group ($F(1,31)=.003$, $p=.957$). The mean number of cognitive operations in Task A was 1024 for the expert group and 1014 for the novice group.
A multivariate analysis of variance with the mean relative frequencies of the seven cognitive operations as dependent variable and group (2 levels) as independent variable showed a significant effect (Wilks' lambda = .380, $F(7,25)=5.825$, $p<.001$). Subsequent univariate tests demonstrated for five operations a significant effect ($\alpha = .05$). In comparison with novices, experts, that is diagnosticians with clinical experience, more often asked for information or gave their opinion (Operation 1)($F(1,31)=11.526$, $p=.002$); they also labelled or abstracted more often(Operation 3)($F(1,31)=5.977$, $p=.020$). Although explaining (Operation 4) appeared rarely in both groups, it appeared more frequently in the expert group($F(1,31)=11.466$, $p=.002$). Neutral matching (Operation 5) appeared most frequently in both groups but was used significantly more by the novices ($F(1,31)=12.964$, $p=.001$). The novices also differentiated (Operation 7) more often than the experts ($F(1,31)=13.750$, $p=.001$).
The two groups were similar in their use of associating (Operation 2) and positive matching or identifying (Operation 6).

Classification of the symptoms in terms of the CBCL syndromes
The average classification score for each item into each girls' syndrome was based on five subjects for both the expert group and the novice group. For the boys' syndromes, this score was based on five subjects for the expert group and four subjects for the novice group. The average classification score for each item into the girls' syndrome 'Aggressive" is given in Appendix B as example.

Table 4.6
Correlation over problem behaviors between the experts' and the novices' mean classification scores.

	Syndrome number*						
	1	2	3	4	5	6	7
boys' syndromes	.77	.69	.71	.75	.51	.12	.58
girls' syndromes	.76	.74	.73	.75	.66	.77	.27

** Boys 1 = agrressive; 2 = delinquent; 3 = depressive; 4 = hyperactive; 5 = socially withrawn; 6 = somatic complaints; 7 = non-communicative. Girls 1 = aggressive; 2 = cruel; 3 = depresssive; 4= hyperactive; 5 = schizoid-obsessive; 6 = socially withdrawn; 7 = somatic complaints.*

The correlations (the Pearson product-moment correlation) between the mean classification scores of the expert group versus the novice group, shown in Table 4.6, turned out to be high (ranging from .51 to .77) except for the boys' Syndrome 6 ($r=.12$) and the girls' Syndrome 7 ($r=.27$), both of which are "somatic complaint" syndromes.

Typicality rating of symptoms

The correlations between the average typicality scores of the expert group versus the novice group turned out to be very high (see Table 4.7).

Table 4.7
Correlations over problem behaviors of the experts' and novices' typicality scores.

	Syndrome number*						
	1	2	3	4	5	6	7
boys' syndromes	84	90	88	87	67	85	83
girls' syndromes	86	91	91	75	85	84	83

**Boys 1 = agrressive, 2 = delinquent, 3 = depressive, 4 = hyperactive, 5 = socially withrawn, 6 = somatic complaints,7=non-communicative Girls 1= aggressive, 2 = cruel, 3= depresssive, 4 = hyperactive, 5 = schizoid-obsessive, 6= socially withdrawn, 7= somatic complaints*

The t-tests on the mean typicality scores for each problem behavior separately showed no more differences in the typicality scores of experts and novices than would be expected by chance (fewer than 5% of these tests showed significant results). This indicates that experts and novices generally agreed in their judgements of the typicality of a certain problem behavior for a syndrome.

Cognitive operations in different classification tasks

The mean total number of operations was considerably higher in task A (1024) than in Task C (267). A univariate analysis of variance confirmed this ($F(1,31)=30.493$, $p<.001$). A multivariate analysis of variance with the mean relative frequencies of the seven operations as dependent variable and task as independent variable also showed a significant effect (Wilks' lambda = .125, $F(7,25)=25.107$, $p<.001$). Univariate tests showed a significant task effect for three operations ($\alpha = .05$). As can be seen in Table 4.5, neutral matching (Operation 5), the Operation that occurred most often in Task A, is the one least often used in Task C ($F(1,31)=77.071$, $p=.000$) and identifying or positive matching (Operation 6) was also used more often in Task A ($F(1,31)=13.571$, $p=.001$). In Task C, labelling (Operation 3) was used considerably more often, in fact it turned out to be the most frequently used Operation ($F(1,31)=103.590$, $p=.000$). No significant differences were found in the occurrence of informing (Operation 1), associating (Operation 2), explaining (Operation 4) and differentiating (Operation 7).

Table 4.9
Overview of the symptoms considered typical (average typicality score ≥6.0, n=24) for each syndrome.

	Syndrome number*														
															Number of typical symptoms
	Boys							Girls							
	1	2	3	4	5	6	7	1	2	3	4	5	6	7	
Items															
1															0
2	6.1			6.0											2
3	8.0	7.4		6.7											3
4	8.0							6.4	6.7						3
5								6.2							1
6				7.3											1
7			7.6		8.0		8.3						7.2		4
8	6.1														1
9	7.2	7.5						6.7							3
10	7.2	7.7							7.0						3
11			7.0	6.4											2
12			7.5				6.2			6.7					3
13	6.7								6.6						2
14		6.2					7.2								2
15		7.7							6.7						2
16															0
17	7.4	6.7						6.5							3
18	7.4							6.0							2
19										6.6					1
20	6.7	8.4							7.9						3
21	6.2	6.7			6.1		6.7								4
22															0
23					6.0	6.0	6.6								3
24	6.5	7.2													2
25				6.8											1
26															0
27	7.9	6.9						7.0	6.9						4
28			8.5		6.7					7.7					3
29							6.7								1
30															0
31												6.3			1
32		6.7													1
33	7.2	6.7						6.7							3
34															0
35							6.5								1
36		7.8							6.6						2
37															0
38			8.3		6.2	7.0				6.8		6.0	6.3		6
39			6.2												1
40	6.2														1
41	6.4														1
42						6.9									1
43															0
44	7.1	6.5													2
45															0
46	7.1	8.1							6.8						3
47			6.5												1
Sum	18	15	7	5	5	3	7	7	8	4	0	2	2	0	83

Boys 1 = agrressive; 2 = delinquent; 3 = depressive; 4 = hyperactive; 5 = socially withrawn; 6 = somatic complaints; 7 = non-communicative. Girls 1 = aggressive; 2 = cruel; 3 = depresssive; 4 = hyperactive; 5 = schizoid- obsessive; 6 = socially withdrawn; 7 = somatic complaints.

Comparing the classifications of the symptoms with the typicality ratings of the symptoms

The correlations between classification scores and typicality ratings ranged from .45 to .76 in the expert group, with a median of .63, and from .26 to .90 in the novice group, with a median of .73 (see Table 4.8). Thus the novice group seems a bit more consistent in their judgement than the expert group.

Table 4.8
Correlations between of the classifications and the typicality ratings of the symptoms for each syndrome.

	Syndrome number*							
	1	2	3	4	5	6	7	median
boys' syndromes								
experts	.76	.73	.72	.66	.56	.45	.68	.68
novices	.73	.75	.82	.75	.67	.42	.68	.73
girls' syndromes								
experts	.62	.53	.58	.65	.52	.72	.52	.58
novices	.73	.67	.90	.72	.76	.76	.26	.73

**Boys 1 = agrressive; 2 = delinquent; 3 = depressive; 4 = hyperactive; 5 = socially withrawn; 6 = somatic complaints; 7 = non-communicative. Girls 1 = aggressive; 2 = cruel; 3 = depresssive; 4 = hyperactive; 5 = schizoid- obsessive; 6 = socially withdrawn; 7 = somatic complaints*

Summarizing the actual embedding of the problem behaviors into the syndromes

To obtain more systematic information on the embedding of the 47 problem behaviors into the predefined CBCL syndromes we computed a summary typicality score for each problem behavior and each syndrome by averaging the experts' and novices' typicality scores. The outcome of the t-tests justified this procedure. Next we defined a symptom as typical for a syndrome if it obtained an average typicality score of 6.0 or more (remember that the summary score can range from 0 to 9) and determined which symptoms were considered typical for which syndromes (see Table 4.9).

From this table it can be seen that 18 symptoms were considered typical for boys' Syndrome 1 (aggressive) and 15 for boys' Syndrome 2 (delinquent). If one compares the number of items considered typical for boys' and girls' syndromes, the difference is remarkable: 60 symptoms typical for the boys' syndromes and 23 for the girls' syndromes.

Nine out of the 47 symptoms were not considered typical for any syndrome. About half of the symptoms (24) were considered typical for more than one syndrome. For example, symptom 38, "easily hurt", was scored as typical for six syndromes. However, one should take into account that the correspondence in items between certain boys' and girls' syndromes is substantial (see Table 4.10).

A minority of the symptoms (14) were rated as typical for only one syndrome.

Table 4.10
Number of corresponding items for the girls' and boys' syndromes

Boys' syndrome number*	Girls' syndrome number* 1	2	3	4	5	6	7
1	19	7	0	4	0	0	0
2	1	6	0	1	0	0	0
3	0	0	8	0	1	0	0
4	1	0	0	8	0	2	1
5	1	2	2	2	0	3	0
6	0	0	0	0	1	0	5
7	1	0	3	0	1	5	0

**Boys 1 = agrressive; 2 = delinquent; 3 = depressive; 4 = hyperactive; 5 = socially withrawn; 6 = somatic complaints; 7 = non-communicative. Girls 1 = aggressive; 2 = cruel; 3 = depresssive; 4 = hyperactive; 5 = schizoid-obsessive; 6 = socially withdrawn; 7 = somatic complaints.*

4.4 Discussion

The aim of this study was to gain insight into the cognitive processes underlying psychodiagnostic classification. This was done with the ultimate goal of looking for heuristic classification rules that could serve as support for an existing empirical classification system.

We identified seven cognitive operations that experienced diagnosticians use in classifying problem behavior. Five of them had discriminating validity in the sense that they were more or less often used by diagnosticians than by novices. Furthermore, we found that the nature of the classification task did have an effect on the classification approach of the expert. The relative frequency of three of the five discriminating operations was comparable in both task conditions. Unfortunately, limited time prevented us from analyzing differences in sequential use of the cognitive operations. Finally, we found that the typicality ratings of the diagnosticians were not different from those of the novices. This last finding may raise some doubt as to the expert status of our experienced diagnosticians. It seems necessary to analyze our somewhat contradictory findings before going on with an attempt to translate the "expert" operations into rules.

Some of our results are compatible with Kolodner's model of the evolution of expertise (1984; Kolodner & Simpson, 1986). Our results suggest that experience has taught the diagnostician to change his reasoning process by questioning, abstracting or labelling the problem behavior more often during classification. This supports the change in reasoning processes that Kolodner's model postulates. It is also consistent with expert-novice shifts that Elio and Scharf (1990) describe in the physics domain. They report a qualitative analysis of the problem statement as being characteristic of an experts approach.

On the other hand, according to Kolodner and Simpson (1986), experience would also reorganize the structure of domain knowledge. Elio and Scharf (1990) also stress the importance of changes in the organization and in the content of

domain-specific knowledge as a result of experience. Our measure of the typicality of problem behavior relates to the cognitive structure in which the domain-specific knowledge is organized. The classification and judgement of typicality reflect conceptions subjects have about the co-occurrence of problem behaviors. As the cognitive structure is supposed to change as a result of experience, we expected a difference in classification and typicality ratings between experienced diagnosticians and novices. Yet no difference was found.

We do not need to doubt the validity of our classification and typicality ratings. In the few empirical studies which have addressed the same task domain of classifying problem behaviors (e.g., Chan & Jackson, 1982; Horowitz, Wright, Lowenstein, & Parad, 1981a; Horowitz, Post, French, Allis, & Siegelman, 1981b), it was also found that the cognitive structure of the more trained or experienced subjects did not differ from that of a less experienced or less trained group. Still, they did find differences between the two groups in terms of the weighting strategy involved in judgment (Chan & Jackson, 1982) and the number of features involved in prototyping (Horowitz et al., 1981a). Therefore, these studies can be seen as instances in which experienced diagnosticians do differ from less experienced persons in processing information about problem behavior, although they do not differ from them with respect to the structure of the domain knowledge.

If we accept Kolodner's model of expert learning, we must conclude that the diagnosticians in this study occupy some intermediate position on the imaginary scale of expertise. Such an intermediate position is in accordance with Hamm's description (1988) of the Dreyfus and Dreyfus theory of expertise, which distinguishes five stages one must go through to become an expert. However, the interesting question then becomes why persons who received a lot of training and do have a lot of experience (some over 20 years) in the clinical psychological domain did not learn enough to become full-fledged experts in the sense of Kolodner's model. In Kolodner and Simpson's (1986) view, expert learning results from the interaction between problem solving on the one hand and experience (including feedback) on the other hand. The feedback that clinicians receive in the mental health field is often flawed; this may account for the fact that so little is learned from experience (Garb, 1989). Insufficient learning from experience accounted for the inaccuracy of expert predictions in the educational field, according to IJzendoorn and Bus (in press). The specialists in this domain are rarely forced to question and correct their diagnostic results.

Furthermore, we believe that an effective interaction only takes place if the person has access to background knowledge in a domain-specific knowledge base. This belief is strengthened by Elio and Scharf's conclusion that: "expert problem-solving depends primarily on having the appropriate domain-specific knowledge and not on any unusual intellectual abilities"(1990, p. 580). We are convinced that marked differences exist in the quality of the domain-specific knowledge bases in different professions. We hypothesize that in the clinical

psychological domain the minimal domain-specific knowledge base that candidate experts should use in order to learn from experience is still in the making.

We do agree with Shanteau (1988) that "The definition of who is an expert is vital to any study of expert decision making" (p. 206). He favours a definition in terms of social competence. The term "expert" refers to those considered by colleagues to be the best at making decisions. Elaborating on ideas about expert learning as presented by Kolodner and Elio and Scharf we emphasize the importance of an appropriate domain-specific knowledge base. If our hypothesis is correct, it would be difficult in the profession of clinical psychology to find persons who are judged to be experts by their colleagues. And if such experts were found, we would expect to find fewer differences in cognitive structures between them and novices than between such experts and novices in a domain such as clinical medicine, where clinicians have access to and are supported by better-developed domain-specific knowledge bases.

4.5 Implications for further research

The results of this study show that experts and novices have the same conceptions about the co-occurrence of problem behavior, which indicates that the prototypes of experts and novices are much alike. This result challenges the expert status of the classification knowledge of experienced diagnosticians. Therefore we considered it important to focus our further research on the value of the clinical classification knowledge as expressed in prototypes. This is be done in the next chapter by comparing the clinical prototypes of experienced diagnosticians with empirical prototypes.

5 Empirical content of intuitive prototypes of experts

5.1 Introduction

Classification of dysfunctional behavior is an important task in clinical decision making (Achenbach, 1985; Blashfield, 1984; De Bruyn, 1985; Rispens, 1986; Rutter, Tuma & Lann, 1988). It focuses or guides, perhaps unintendedly, the generation of hypotheses concerning the conditions which elicit or sustain the identified problem. It narrows down the field in terms of causes, treatment and prognosis and, according to Rutter (1975), it provides a shorthand language for communication with other professionals.

In the study reported in the previous chapter, we focused on the classification behavior of experienced diagnosticians. In this chapter we will focus on a different aspect: intuitive prototypes, i.e., the classification knowledge incorporated in cognitive structures of experienced diagnosticians. The sample of diagnosticians in the study of chapter four was fairly representative, and most of the subjects had some knowledge of the CBCL-system (Achenbach & Edelbrock, 1983), the classification system with which we confronted them in that study. Nevertheless, it appeared that only one of the 22 used the CBCL-system in practice. The diagnosticians seemed to rely during the classification task on their experience (as recorded in cognitive structures or memory). (Of course the instructions for the tasks stimulated them to use knowledge obtained from experience.) In view of the biases that strongly influence cognitive processes, it can be questioned whether it is advisable to rely so strongly on memory when one has to classify problem behaviors.

Research on intuitive prototypes in psychopathology has demonstrated that the prototype view is in correspondence with the way how clinicians actually think about and use diagnostic categories (Cantor, Smith, French & Mezzich, 1980). Prototypes (whether "intuitive" or "empirical") can be regarded as classes or categories in which features are organized. In the prototype view of classification, category membership is a matter of degree. In the domain of child psychopathology this means that problem behaviors can be more or less typical for a certain prototype.

There are different viewpoints as to the value of these intuitive prototype. Chan and Jackson assert that "to the extent that implicit personality theory [comparable with intuitive prototypes] arises from the observation of co-occurrence of behaviors in people, it reflects the distribution of traits in people and provides a relatively valid foundation for the judgement of specific targets" (1982, p.5). However, Chapman and Chapman (1982) proved that subjects see a correlation between two events as being stronger than it actually is. So in view of this "illusory correlation" it is quite possible that intuitive prototypes do not reflect the structure of reality, but contain sets of features that people think of as occurring together. The issue of whether these cognitive structures or prototypes reflect the structure of reality has been raised (Horowitz, et al., 1981a; Rosch, Mervis, Gray, Johnson & Boyes-Bream, 1976). However, we know of no study

that compared the intuitive prototypes of clinicians with the structure of reality or determined the empirical content of these prototypes.

The present study addresses the value of intuitive prototypes by comparing them with empirical prototypes. By intuitive we mean that the structure is acquired through the practice of the art, through the classification of subjectively selected features. By empirical we mean that the structure is acquired through multivariate analysis of systematically collected features.

We assume that experience, i.e. the observation and classification of problem behaviors in children, has created a cognitive structure in the memory of the diagnostician that includes impressions of the frequency of co-occurrence of problem behaviors. If two behaviors are assumed to co-occur relatively frequently, then when either one is given, the other one is more likely to be expected. These cognitive structures can be seen as intuitive prototypes, i.e., the categories or classes of the implicit classification knowledge of the clinician.

Horowitz, Wright, Lowenstein, and Parad (1981a) describe a method for obtaining intuitive prototypes of clinicians. They asked clinicians to list the most common characteristics that they think of when they imagine, for instance, a typical aggressive-impulsive child. They viewed this composite of features as a cognitive structure that depicts an idealized member of that category. The method of presenting stimuli and listing associated features or attributes is commonly used in basic research on prototypes (see e.g., Rosch et al., 1976) and also in research on psychopathological prototypes (Cantor et al., 1980; Chan & Jackson, 1982). For example, Cantor et al. (1980) asked clinicians to list clinical features that characterize the prototypical patient.

In this study we address the following questions: a) What is the correspondence between an empirical and a clinical, intuitive ranking of symptoms in syndromes? b) What is the correspondence between an empirical and a clinical, intuitive clustering of symptoms? c) Do the same problem behaviors that clinicians think of as co-occurring within an incompletely specified syndrome also co-occur empirically?

5.2 Method

Design and subjects

The study involved the same two groups of experts that took part in the study on diagnostic classification reported in the previous chapter.

The design of the study in figure 5.1 shows the assignment of the tasks to the subject groups.

Tasks and procedure

We developed three tasks assumed to invoke intuitive prototypes of the subjects, in particular the notions diagnosticians have of the co-occurrence of problem

behaviors. These prototypes contain features-in the case of our study problem behaviors-that are more or less typical for that prototype or diagnostic category.

For the construction of the tasks we made use of the empirical classification system developed by Achenbach and adapted for a Dutch population by Verhulst (1985; Verhulst, Koot, Akkerhuis & Veerman, 1990). We used the items of the CBCL, a standardized format to record the behavioral problems and competencies of children as reported by their parents. We used the 7 empirical syndromes for boys aged 6-11 and the 7 empirical syndromes for girls aged 6-11 that were found in the Dutch population (Achenbach, Verhulst, Baron & Althaus, 1987a; Verhulst, Achenbach, Althaus & Akkerhuis, 1988). In addition, we used the core syndromes, i.e., clusters of co-occurring problem behaviors that are found in different age and gender groups and thus reflect a very steady empirical grouping of problem behaviors (Achenbach, Conners, Quay, Verhulst & Howell, 1989; Koot & Verhulst, 1990). In Appendices A and C complete descriptions of the empirical syndromes for boys and girls aged 6-11 and the core syndromes are given.

	Task 1 rank	2 cluster	3 complement	N
expert group 1	x			(n=12)
expert group 2		x	x	(n=10)

Figure 5.1 The design of the study 1: Ranking of symptoms in existing empirical syndromes; 2: Intuitive clustering of the original CBCL symptoms; 3: Complementing incompletely specified syndroms.

Task 1: Ranking of symptoms in existing empirical syndromes

In Task 1, presented to the first expert group, we asked subjects to rank the items of an existing empirical syndrome. Each syndrome was put on a card with the defining items printed in a random order. The labels of the syndromes were not included on the cards. There were seven cards representing the syndromes of the boys and seven cards representing the syndromes of the girls. To facilitate ranking performance, the items of each syndrome were also printed separately on small cards. The subjects received the seven syndrome cards of one gender group one by one, in a random order, and the matching set of small cards. They were asked to rank the items in the syndrome according to the typicality of the items for that particular syndrome according to their experience, starting with the most typical item. The outcomes of Task 1 provide evidence on the correspondence between an ordering based on clinical experience and an ordering based on empirical data.

Task 2: Intuitive clustering of the original CBCL symptoms

A description of this task has already been given in chapter 4, but it is repeated here for the sake of convenience.

In Task 2, diagnosticians were asked to cluster the original items of the CBCL

according to their own points of view (9 subjects were not familiar with the CBCL). Subjects were first given a set of 112 small cards containing the problem behaviors and child characteristics as used by the CBCL. From this set they could select the behaviors and child characteristics with which they were familiar. Problem behaviors and characteristics which they had encountered in their practice but which were not in the CBCL set could be written down on new cards. Next, the subjects were asked to sort the collected behaviors and characteristics into clusters containing items that in their own clinical experience co-occurred frequently. They were allowed to form as many clusters as they wanted and to put the same item into more than one cluster. In this task, we focused on the question of whether the outcomes of an intuitive clustering are in agreement with an empirically established system of clusters as represented in the core syndromes.

Task 3: Complementing incompletely specified syndromes
In Task 3, diagnosticians were asked to complement seven incompletely specified syndromes presented on vignettes. Each vignette contained one randomly chosen third of the symptoms of a core syndrome of the CBCL. To imitate clinical practice, we put a name on the vignette, as illustrated in the example below.

Yvonne refuses to talk and is shy and timid. She is unhappy and sad.

The seven vignettes were given to the subjects on separate pieces of paper one by one in random order. We asked the diagnostician to complement the artificial case description by reporting which additional problem behaviors they expected to be present in that particular child. The outcome of this task provides information on which behaviors are judged to co-occur with the behaviors listed in an incomplete description of an empirical core syndrome. We investigate whether these behaviors are similar to the behaviors that empirically co-occur.

The tasks were administered in the same session(s) as the tasks of a previous reported study (see Chapter 4). For Task 1 no time limit was set. All subjects completed this task in less than 30 minutes. For Task 2, a one-hour time limit was agreed upon and for Task 3 five minutes for each case description. Tasks 2 and 3 were administered in one session for most subjects. Task 2 was always given before Task 3. For two diagnosticians Task 3 was administered in another session because of lack of time. The sessions with the diagnosticians took place at their own offices, while the sessions with the novices took place at the University of Nijmegen. Task 1 was administered by the author, Tasks 2 and 3 by two assistants.

Data Analysis
The data were mostly analyzed in a qualitative way. This section contains a detailed description of the procedures used to analyze the data and is fairly

extensive. We often refer to the Appendices to illustrate the procedures. For the analysis of the data we used the empirical classification knowledge obtained with the CBCL in a Dutch population (Koot & Verhulst, 1990; Verhulst, Koot, Akkerhuis & Veerman, 1990).

Analysis of the ranking of symptoms

For each subject, rank correlations (Spearman's rho) were computed between the empirical order of the items in each syndrome and the clinical order of the items. The factor loading of the items, as found by Verhulst in a Dutch population, determined the ranks of the items in the syndromes and provided the empirical order (see Appendix A). The ranks experts gave to the items in task one determined the clinical order of the items in each syndrome. In order to examine the possibility that some syndromes are better known than others, we counted the number of significant correlations (α=.10) for each syndrome.

Analysis of intuitive clustering

We first counted the number of items not selected from the original CBCL itempool, (representing problem behaviors clinicians were not familiar with), and the number of items added to the CBCL itempool. Next we counted how many intuitive clusters had been formed and the number of items in each cluster. The resemblance between the intuitive clusters and the core syndromes was computed by the following procedure, which is best clarified by an example (see Table 5.1).

Table 5.1
Illustration of method for computing the similarity between an intuitive cluster of one diagnostician and the core syndromes

	Core syndromes									
	externalizing		mixed			internalizing		boys only	girls only	
number of items	1 (19)	4 (11)	3 (9)	5 (5)	6 (2)	2 (13)	7 (9)	8 (4)	9 (5)	rest
intuitive cluster neurotic (17)	1 5.9%	0	0	0	0	9 52.9%	4 23.5%	0	0	3 17.6%

1 = aggressive; 2 = anxious depressed; 3 = attention problems; 4 = delinquent; 5 = schizoid; 6 = physical complaints; 7 = withdrawn; 8 = socially inept (boys only); 9 = cruel (girls only).

Suppose a diagnostician has made a cluster labelled "neurotic" consisting of 17 items. Suppose further that 1 out of 17 behaviors (or 5.9%) occurs in Core Syndrome 1 "aggressive"; 9 out of 17 behaviors (or 52.9%) occur in Core Syndrome 2 "anxious-depressed"; 4 behaviors (or 23.5%) occur in Core Syndrome 7 " withdrawn"; 3 behaviors (or 17.6%) occur in no Core Syndrome. This intuitive cluster has the greatest overlap with the empirical core syndrome "anxious-depressed". For each core syndrome, the percentage of symptoms in the intuitive cluster that fall into that core syndrome indicates the resemblance of the

intuitive cluster to the core syndrome. When all items of an intuitive cluster are part of a single core syndrome, there is just one single percentage (namely 100%) representing the resemblance. Note that this does not imply that the intuitive cluster is exactly the same as the core syndrome since the core syndromes can still contain more items than the ones present in the intuitive cluster.

Analysis of the complements of incompletely specified syndromes

The correspondence between the behaviors that the subjects judged to be co-occurring with incomplete syndromes and the behaviors that empirically co-occur was computed as follows: First two raters separately determined whether the complemented behaviors of the diagnosticians were synonymous with CBCL items. In case of disagreement the author of the present study was consulted. It was not required that the complemented behaviors were literally the same as the CBCL items. From the complemented behaviors we deleted the behaviors that were added more than once.

For each vignette we then determined the number of added behaviors that were similar to the CBCL items that empirically complement the core syndromes. This number of "correctly added" behaviors was divided by the total number of items that empirically complement the core syndrome. So the resulting number indicates the proportion of empirically established complementary behaviors that is in fact recovered or recognized by the clinician. Again we provide an example to illustrate the procedure (see Table 5.2).

Table 5.2
Illustration of procedure for computing the proportion of empirically complementary items uncovered by diagnosticians

Step	To be counted	Number
1	All items complemented by the diagnostician	5
2	Items synonymous with a CBCL item	3
3	Items complemented by the diagnostician more than once	0
4	"Correctly" complemented items	1 (=a)
5	Items that empirically complement core syndrome seven	6 (=b)
Proportion of "correctly" complemented items=a/b		1/6=16.7%

Suppose a diagnostician has complemented core syndrome 7 which consists of nine items, three of which were presented to the subject in the vignette, with five behaviors. Three behaviors are synonymous with a CBCL item. Only one item is identical to an item that empirically complements the core syndrome. In all six CBCL items complement the core syndrome empirically. The similarity between the behaviors complemented by the diagnostician and the items that empirically complement the core syndrome is 1/6, i.e. 16.7%.

5.3 Results

Ranking of symptoms in existing empirical syndromes

The rank correlations (Spearman's rho) between the empirical order and the clinical order of the items in each syndrome turned out to vary considerably over subjects (see Table 5.3).

Table 5.3
Rank correlations (Spearman's rho) between the empirical order and the clinical order of items in a syndrome

	boys' syndromes						
	1	2	3	4	5	6	7
group							
expert							
2	.33	.24	.32	.41	-.14	.13	.61**
4	-.12	.00	.35	.36	-.55	.32	.36
6	.26	-.09	.55**	.04	-.02	.39	.24
8	-.11	.06	.73**	.55**	-.26	.79**	.39
10	.50**	-.06	.85**	.52*	.36	-.07	.71**
12	.51**	.04	.40	.70**	.57*	-.29	.54**
	girls' syndromes						
	1	2	3	4	5	6	7
group							
expert							
1	.14	-.27	.42**	.27	.04	.52*	.43
3	.23	.78**	-.45	.47*	.44*	.12	.50
5	.41**	.61**	.04	.42*	-.20	.40	.71**
7	-.40	.27	-.09	.34	-.10	-.33	.32
9	.38**	.52**	.20	.28	.20	.63**	.79**
11	.00	-.02	.13	.25	.22	.28	.43

**p<.05;* p<.10

Boys 1 = agrressive; 2 = delinquent; 3 = depressive; 4 = hyperactive; 5 = socially withrawn; 6 = somatic complaints; 7 = non-communicative. Girls 1 = aggressive; 2 = cruel; 3 = depresssive; 4 = hyperactive; 5 = schizoid-obsessive; 6 = socially withdrawn; 7 = somatic complaints.

For each expert it was possible to have 7 significant rank correlations (at $\alpha = .10$). Since there were 12 subjects (diagnosticians) the maximum number of significant rank correlations is 7 x 12 = 84. The data show that in the expert group 28 out of 84 (or 33%) of the orderings correspond significantly to the empirical ordering.

In Table 5.3 it can be seen that the following syndromes were ordered relatively often in correspondence with the empirical rankings: Syndrome 1, "aggressive" and Syndrome 4 "hyperactive", for both boys and girls; Syndrome 2 for the girls, "cruel", and Syndrome 3, "depressive", and Syndrome 7, "non communicative" for the boys.

Intuitive clustering of the original CBCL-symptoms

Table 5.4 shows that the number of items not selected from the original item pool of the CBCL ranges from 1 to 19. The number of items that were added ranges from 0 to 11. Appendix D and E report which items were omitted and which were added.

Table 5.4
Number of items not selected from and added to the original CBCL item pool

	Number of items		
	Omitted	Added	In final item pool
Subject			
1	19	7	100
2	5	0	107
3	19	6	99
4	17	0	95
5	1	4	115
6	11	3	104
7	1	11	122
8	4	2	110
9	7	3	108
10	3	0	109
Median	8.7	3.6	106.9

The size of the final itempool the diagnosticians used for the construction of their clusters varied from 95 to 122 items.

On the average a diagnostician formed 12 clusters (see Table 5.5).

Table 5.5
Overview of the intuitive clustering (Task 2).

	Number of clusters	Total number of items used	Mean number of items per per cluster	Number of items used more than once
Subject				
1	4	104	25.8	4
2	12	107	8.9	0
3	11	188	17.1	89
4	12	140	11.7	45
5	11	114	10.5	0
6	16	111	6.9	7
7	17	212	12.5	90
8	7	193	27.6	83
9	9	126	14	18
10	22	199	19	90
Mean	12.1	149.4	15.4	42.6

The diagnosticians on the average formed more clusters than were empirically found by Koot and Verhulst. The empirical clustering consists of 9 clusters (as well as a miscellaneous group): seven core syndromes, one syndrome found only in boys, one syndrome found only in girls. Moreover, the clusters of the diagnosticians contain on the average more items (15) than the core syndromes (10).

The labels the diagnosticians gave their clusters are shown in Appendix F. In Appendix G we present percentages of overlap for all intuive clusters. All in all, 121 intuitive clusters were compared with 9 core syndromes, resulting in 9x121=1089 comparisons. If one looks at the percentages of overlap one notices that most intuitive clusters have an overlap with more than one empirical cluster. The intuitive clusters for which a 100% overlap is reached consist of only one or two items. In very few instances (19 out of 1089) an intuitive cluster contains more than 50% of the items of one core syndrome (see Table 5.6).

Table 5.6
Number of times that the overlap between an intuitive cluster and a core syndrome included more than 50%, for each of the 10 diagnosticians

	Core syndromes									
	externalizing		mixed			internalizing		boys only	girls only	
number of items	1 (19)	4 (11)	3 (9)	5 (5)	6 (2)	2 (13)	7 (9)	8 (4)	9 (5)	sum
Subject										
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	1	0	0	0	1
3	0	0	0	1	0	0	0	0	0	1
4	0	0	0	0	0	0	0	0	0	0
5	1	0	1	0	0	1	0	0	0	3
6	1	0	0	1	0	0	0	0	0	2
7	1	0	3	0	0	0	0	0	0	4
8	0	0	0	0	0	0	1	0	0	1
9	1	0	0	0	0	0	0	0	0	1
10	0	0	1	0	0	0	0	5	0	6
Sum	4	0	5	2	0	2	1	5	0	19

1 = aggressive; 2 = anxious depressed; 3 = attention problems; 4 = delinquent; 5 = schizoid; 6 = physical complaints; 7 = withdrawn; 8 = socially inept (boys only); 9 = cruel (girls only).

So these findings show that intuitive clusters contain items from different core syndromes. Parts of the empirical groupings of a core syndrome can be found in the intuitive clusters, where they are combined with parts of other core syndromes.

Complementing incompletely specified syndromes

The percentages indicating the extent to which the diagnosticians recovered the items that empirically complement the core syndrome are small (see Table 5.7). Two diagnosticians (9 and 10) were able to complement core Syndrome 5, "schizoid", with all of the empirically co-occurring behaviors.

The recovery percentages averaged over the seven syndromes range from 1.1 percent to 25.7 percent. Four diagnosticians scored just above 20 percent.

Table 5.7
Similarity between the complemented behaviors of the diagnosticians and the items that empirically complemented the core syndrome.

	Core syndrome 1	2	3	4	5	6	7	
	Number of complements (13)	(9)	(6)	(7)	(3)	(4)	(6)	Mean
Complement of subject								
1	0	0	16.7	0	33.3	0	16.7	9.5
2	7.7	11.1	33.3	0	33.3	0	0	12.2
3	7.7	11.1	50	42.9	0	0	33.3	20.7
4	23.1	11.1	16.7	0	0	0	16.7	9.6
5	23.1	22.2	16.7	42.9	0	25	50	25.7
6	7.7	11.1	16.7	0	0	0	0	5.1
7	7.7	0	0	0	0	0	0	1.1
8	0	11.1	16.7	14.3	0	25	16.7	12
9	0	0	50	0	100	0	0	21.4
10	7.7	22.2	16.7	0	100	0	0	20.9

1 = aggressive; 2 = anxious depressed; 3 = attention problems; 4 = delinquent; 5 = schizoid; 6 = physical complaints; 7 = withdrawn; 8 = socially inept (boys only); 9 = cruel (girls only).

Note: Each entry specifies the percentage of remaining behaviors in the core syndrome that the diagnostician complemented.

5.4 Discussion

The aim of this study was to examine the intuitive classification structures of experienced diagnosticians, in particular the empirical content of the intuitive prototypes. This was done by comparing intuitive prototypes from diagnosticians with empirical prototypes obtained through multivariate analysis. We found that 33% of the intuitive rankings of behaviors in existing syndromes corresponded with the empirical rankings based on factor loadings. Specifically Syndrome 1, "aggressive", Syndrome 4, "hyperactive" for both boys and girls; Syndrome 2, "cruel" of the girls, and Syndromes 3, "depressive", and 7, "non-communicative" of the boys, were ordered more often in correspondence with the empirical rankings. Furthermore, we found that the intuitive clusters made by the diagnosticians corresponded very poorly with the empirical core syndromes. It can be concluded that the intuitive cluster consists of parts from different core syndromes. Finally, we found a low similarity between behaviors that clinicians judge as co-occurring with an incompletely specified syndrome and behaviors that empirically co-occur with that same syndrome. In summary, we therefore conclude that intuitive prototypes do not correspond with empirical prototypes -in other words, that the empirical content of intuitive prototypes is minimal.

Our study is truly explorative in the sense that we cannot compare our results with other research findings. As mentioned in the introduction to this chapter, we do not know of any study that has investigated the empirical content of intuitive

prototypes. Before interpreting our findings, we consider it necessary to discuss some aspects of the method used.

First of all, one could question whether the tasks called into play the implicit classification structures diagnosticians use in daily practice. It is difficult to answer this question, because all three tasks focused on "intuitive" classification knowledge, namely the notions diagnosticians have about the co-occurrences of problem behaviors as represented in prototypes. As demonstrated by Cantor et al.(1980), the prototypical view of classification is in correspondence with the way clinicians actually think about and use diagnostic categories. Diagnosticians are often not aware which classification structures they use or have in mind (Van Mechelen & De Boeck, 1989). Some even suppress the idea of using classification structures at all. The tasks we constructed do not reflect a conscious activity during the diagnostic process. Usually, clinicians do not consciously rank problem behaviors during clinical assessment as we requested them do to in Task 1, or verbalize their classification structures as we requested in Task 2. These tasks may therefore to some extent be artificial for the clinicians. In Task 3, however, we tried to imitate clinical practice by presenting a short case description. This task was based on the assumption that, if a clinician has some information about certain problem behaviors of a child, he will also have notions about the co-occurrence of other problem behaviors the child will have. By presenting certain problem behaviors we expected to uncover these notions.

In our study we actually used three different tasks that called into play intuitive prototypes of clinicians. The fact that all three tasks showed similar findings, namely minimal correspondence between intuitive and empirical prototypes, gives us confidence in the replicability of our results. We therefore predict that future studies on this topic will show similar findings.

Now, of course, one can object to the use of the empirical structures as a standard of comparison. As far as the ranking task is concerned, we realize that perhaps we have been too severe. We expected clinicians to produce a ranking similar to the empirical ranking. However, the factor loading of an item or symptom, on which the empirical ranking was based, is not intrinsically stable. Achenbach himself does not rank the items in a syndrome (personal communication, June 1990). In view of this, we decided to liberalize our criterion afterwards and checked whether the 50% most typical items for each syndrome according to the clinicians' judgement were similar to the 50% most typical items determined empirically. The correspondence between the intuitive and empirical ranking, as expressed in the average percentage agreement or similarity, became 63%. Although the correspondence seems to have increased in comparison with the aforementioned 33%, we must take into account that a great deal of this similarity could occur by chance. Only 14 out of 84 kappa 's, or 17%, turned out to be statistically significant.

In the other tasks we used the core syndromes as a comparison criterion. These core syndromes can be seen as very solid empirical knowledge that best reflects the

reality of co-occurrences of problem behaviors. But for these tasks the difference between the intuitive prototypes and empirical prototypes also remained.

Why do intuitive prototypes deviate from the structure of reality as reflected in the core syndromes? In our opinion two factors can account for this. As it is assumed that intuitive prototypes or implicit classification structures are based on the *observation* of co-occurences of problem behaviors, the clinical population with which a clinician is confronted in his or her practice then becomes an important factor for the construction of these prototypes. The population from which the empirical syndromes were derived is probably much larger and more diverse than the clinical population on which the clinician has constructed his prototypes. The setting in which a clinician is working will very much influence the observation of co-occurences of problem behaviors and thus the construction of intuitive prototypes. A diagnostician who works in a child psychiatric setting will deal with a different group of children and will observe different co-occurrences of problem behaviors than a diagnostician who works in a general mental health agency. As a consequence these two persons will have developed different prototypes and it is very well possible that they would group the same problem behaviors of a child in different clusters. An explanation that resembles the previous one is given by Rosch et al. : "One evident aspect of expertise is that the expert's knowledge is probably often confined to specific parts of a taxonomy, thereby creating unevenness in the expert's categorization of that taxonomy" (1976, p.432). Someone who is specialized in the assessment of depressive disorders in children will have developed a different prototype than someone not specialized in depression.The specialist will probably have more differentiations within the category depression. The role of "expertise" in the development of intuitive prototypes in child psychopathology and the consequences for the judgement as to whether a child with certain problem behaviors does or doesn't have a certain disorder, needs further examination. It would also be interesting to examine if prototypes of clinicians working in the same setting are more alike.

Another important factor that can account for the deviation between intuitive prototypes and the structure of reality is the fact that observational capacity of clinicians is biased by phenomena such as "illusory correlation" (Chapman and Chapman, 1982) and the "inability to assess covariation" (Achenbach, 1985).

The apparent differences between intuitive prototypes and the structure of reality would make it necessary to view critically one characteristic of experts, i.e., the confidence they usually have in their judgements (Shanteau, 1988; in press).

In view of the results of this study we are obliged to advise clinicians to use the available empirical knowledge on the co-occurrence of problem behaviors. Today no clinician would judge a child's intelligence profile using his own experience-based implicit classification structure, instead he or she uses an empirically validated test for it, e.g. the WISC-R. Yet, in clinical practice the very same clinicians rely heavily on their own possibly biased memory to judge a child's psychopathology, even though a more standardized method is available.

6 Summary and implications

In this final chapter we will first summarize the main findings of this thesis. Next we will discuss some implications for the notion of an expert and expert knowledge in the psychodiagnostic domain, in particular in relation to the development of expert systems. Finally we will make some suggestions for future research and for diagnostic practice.

6.1 Summary of findings

This thesis is concerned with expert knowledge with a view to developing an expert system as a decision aid for psychodiagnostics. We focused on one important part of the diagnostic process: the classification of problem behaviors.

In Chapter 2 we analyzed the role of classification in the diagnostic process and the use of classification systems to assist the clinician during this task. We discussed two important classification approaches in the field of child psychopathology, the DSM-III-R system and the CBCL-system, and we concluded that the empirically based CBCL system appears to be the most promising one in view of our purposes. We questioned whether the CBCL is sufficiently representative to be used as a data base in a computerized diagnostic knowledge system. If certain problem behaviors are not represented in the CBCL they will not be detected in a forthcoming case. In Chapter 3 we report an empirical study into the content validity of the CBCL. The study demonstrated that the items of the CBCL do not cover the broad range of problem behaviors reported in diagnostic reports. We concluded that several items should be added. These findings raised the question of how to decide whether the added items or symptoms are part of the empirically derived syndromes and illustrates a functional problem in the use of classification systems: how to deal with individual problem behavior that doesn't fit into the classes of a system. Following Kleinmuntz, who was able to extract the decision rules an expert was using in the classification of MMPI-profiles, we focused our research on the classification process of experienced diagnosticians in order to obtain practical guidelines for dealing with such a classification problem. In Chapter 4 we present the study on how experienced diagnosticians, referred to as experts, classify individual problem behavior, in comparison with novices. Inspired by Kolodner's model of the evolution of expertise, which assumes that experience reorganizes the structures of both the reasoning process and the domain knowledge, we examined the differences between experts and novices. We identified seven cognitive operations that experienced diagnosticians use in classifying problem behavior: (a) asking or giving information, (b) associating, (c) abstracting or labelling, (d) explaining, (e) neutral matching, (f) identifying, and (g) differentiating. Results showed that five of them were more or less often used by experienced diagnosticians than by novices. This finding suggests that experience has taught the diagnostician to change his reasoning process by questioning,

abstracting or labelling the problem behavior more often during classification and supports the change in the reasoning process that Kolodner's model postulates. But we also found that the outcomes of the classification and the typicality ratings given by the experts were not different from those of the novices. Experts and novices have the same conceptions about the co-occurrence of problem behavior, i.e., have similar prototypes. This last finding disagrees with the change in domain knowledge Kolodner's model hypothesizes and casts doubt on the expert status of experienced diagnosticians. We subsequently focused our research on an evaluation of intuitive prototypes, i.e., the classification knowledge as represented in cognitive structures of experienced diagnosticians. In Chapter 5 we examined the intuitive classification structures of experienced diagnosticians, in particular the empirical content of the intuitive prototypes. We compared intuitive prototypes of diagnosticians with empirical prototypes derived by multivariate analysis. Results showed that the intuitive clusters made by the diagnostician correspond very poorly with the empirical core syndromes, but rather consist of parts from different core syndromes. The similarity between behaviors that clinicians judge as co-occurring with an incompletely specified syndrome and behaviors that empirically co-occur with that same syndrome was likewise low. We concluded that intuitive prototypes do not correspond with empirical prototypes, in other words, that the empirical content of intuitive prototypes is minimal.

6.2 The notion of an expert and the development of an expert system.

In this thesis we were confronted with the difficulty of defining "expert" and "expertise" in the domain of psychodiagnostics. In general, an expert is someone who knows the most about a particular domain and is considered best in solving a particular problem typical for that domain. He or she has acquired this knowledge or ability by the practice of the art, by experience. Since objective standards for evaluating the outcome of the diagnostic process are lacking, it is important to evaluate critically the knowledge of the diagnosticians, or their expert status (Carroll, 1987). In our study we used two different criteria to evaluate the private classification knowledge of experienced diagnosticians. In Chapter 4 we used the experience criterion, or expert-novice differences, and in Chapter 5 we used an empirical criterion, i.e. the empirical classification knowledge as provided by the CBCL. The results of these studies cast doubt on the expert status of diagnosticians.

The correspondence between the classification outcomes of experts and novices does not support a change in basic domain knowledge as a result of experience (i.e., dealing with cases); it shows that the diagnosticians in our domain are not full-fledged experts in the sense of Kolodner's model. The main finding of Chapter 5 -the minimal empirical content of the intuitive classification

structures- makes one wonder if it is correct to consider someone an expert whose classification structures do not reflect the structure of reality.

Can we conclude from these results that there are no experts in the psychodiagnostic domain?

First of all it is important to note that the study in this thesis addressed just one aspect of psychodiagnostics, the classification of problem behaviors . The results of other studies on expert-novice differences (Garb, 1989) as well as studies on the validity of prognoses (IJzendoorn & Bus, in press) do support a generalization of results from this study to other parts of the diagnostic process.

However, it can be argued that the selection of experienced diagnosticians as potential experts does not allow the aforementioned conclusion that there are no real experts. Experts are those who have reached the pinnacle of their profession, according to Shanteau, and they can be found in the top of a pyramid framework (1988). Our selection of diagnosticians was not pointed at the top of this pyramid but at the finding of experienced diagnosticians representative for the different work settings in the field. The results of our study thus can be generalized to the population of experienced diagnosticians, but they do not sufficiently justify the conclusion that there are no experts in the psychodiagnostic domain. (Only that we didn't find any in our sample.)

Furthermore, instead of concluding that there are no experts in the psychodiagnostic domain, we suggest expanding the notion of an expert with a new aspect, i.e. the methodological correctness of his or her diagnostic procedure. At this moment, the lack of objective standards for validating the outcome of the diagnostic process and the lack of scientific knowledge in child psychopathology seems to demand an explicit or transparent diagnostic procedure (for instance the normative diagnostic cycle model as described in chapter two) to evaluate the expertise of the clinician. We therefore advise incorporating a methodological criterion in the notion of an expert and assume that the experienced diagnosticians will gain from the application of an accurate diagnostic procedure to real cases, which will (it may be hoped) result in expertise. As yet we consider the use of a correct diagnostic procedure a necessary but not sufficient condition for the development of expertise in the psychodiagnostic domain. Experience, the dealing with the cases, and the use of adequate substantial domain knowledge (empirically and theoretically valid) are the other conditions necessary for the development of expertise.

Very closely related to the notion of an expert is the idea of mimicking the expert diagnostician in a computer program, i.e., of developing an expert system. In the introduction it was noticed that an expert system as a decision aid is supposed to improve the quality of current psychodiagnostic practice. However, in view of the poor performance of clinicians found in the present study and other studies in clinical psychology (Shanteau, 1988; in press) it is not likely that an expert system based on a descriptive model of clinical judgement will improve the quality of the

diagnostic practice. Results of this and other studies (Garb, 1989) show that in psychodiagnostics the knowledge gained from experience is not the same as expertise and thus it is not profitable to base an expert system on the knowledge of experienced diagnosticians.

The question then becomes whether we should skip the idea of developing an expert system and focus on the development of other decision aids. However, the combination of the large amount of information that has to be processed in the diagnostic task with the concept of bounded or limited rationality (cf. Newell and Simon, 1972), seems to justify the use of a computerized diagnostic decision aid for the diagnostic practice. If we want such a computerized aid or system to be considered an expert system, its content should meet the same criteria that we applied to the notion of an expert. This means that the system should contain (a) a methodologically correct diagnostic procedure, for instance the normative diagnostic cycle, (b) adequate substantial domain knowledge, for instance the CBCL system, and (c) a reasoning method that learns from experience, for instance case-based reasoning.

In case-based reasoning (CBR) a problem is solved by matching the problem description to a previously solved case, using the past solution in solving the new problem. In CBR, learning occurs as by-product after each problem solving session, in that relevant information from a problem just solved is retained, making the new experience available for future problem solving. CBR is a very common problem solving method that is also applied in medical diagnosis (Aamodt, 1990). One of the requirements Aamodt claims for future knowledge-based systems is "a problem solving and reasoning method that is able to effectively combine reasoning from past cases with reasoning within a competent and robust model of more general domain knowledge (1990, p.2)".

As far as the task of classification of problem behaviors is concerned, this implies that the knowledge-based system or expert system should contain empirical classification knowledge as provided by the CBCL in combination with a case-based reasoning method to deal with the problem behaviors not included in the classification system.

Kolodner and Simpson describe an attempt to implement the case-based reasoning method in a computer system for psychiatric diagnosis, called SHRINK. This system contains the domain knowledge as provided by the DSM-III and when complete "will do both diagnosis and treatment of affective (mood) disorders based on analogy to previous cases and will also track down and explain its failures, correcting the knowledge that was at fault" (1986, p.113).

We think it is worthwhile to focus research on the possibility of the implementation of the case-based reasoning method in a computerized diagnostic decision system for the domain of problem behaviors in children.

6.3 Suggestions for further research and diagnostic practice.

In view of the disappointing results in studies on expert-novice differences, one can question whether research should continue to search for expertise in the field of psychodiagnostics. Or one can argue that until now research did not succeed in retrieving the expertise of clinicians because the studies didn't use the proper task, or selected the wrong subjects, or used an inappropriate elicitation method.

As far as the task is concerned, it is of course possible that the expertise of diagnosticians will show up in other tasks within the diagnostic process, e.g., during the generation or testing of hypotheses, or during treatment recommendation. However, this assumption doesn't seem very likely in view of the biases that can occur during these tasks, e.g. ignoring base-rates, unwillingness to entertain alternative hypotheses (see Kahneman, Slovic, & Tversky, 1982; Nurcombe & Fitzhenry, 1986) and, related to it, in view of the development of decision aids to deal with these biases, e.g. the use of Theorem of Bayes for hypothesis evaluation in psychodiagnostics (Wouters, 1989).

In view of our findings and the discussion concerning the notion of an expert we suggest selecting diagnosticians more carefully in future studies that aim to find expertise, and to include the methodological criterion. From the finding that neuropsychologists who were reputed to be experts made more valid ratings (Garb, 1989) it appears that the use of a social competence definition, as suggested by Shanteau (1988), can enhance the likelihood to finding real experts and expertise.

With respect to elicitation methods, we note that up to now very little is known about the effect of the elicitation method on the knowledge extracted from experts. Relevant research is now being carried out (Brinkman, 1990; Burton, Shadbolt, Rugg, & Hedgecock, 1990) and in view of the time-consuming aspects of some methods, e.g. protocol analysis, it certainly can be considered important to investigate this potential method effect further.

In Chapter 5, it was concluded that the effect of expertise on intuitive classification knowledge should be examined in more detail in the future. Rosch, Mervis, Wayne, Johnson and Boyes-Bream (1976) have stated that expertise creates unevenness in a taxonomy. As yet we do not know what this unevenness in taxonomy might consist of, e.g., in a specialist in depression; or on the classification judgement by that specialist. Therefore we strongly advise clinicians and candidate clinicians to use the available empirical classification knowledge on the co-occurrence of problem behavior and not to rely on their own possibly biased memory.

This thesis addressed the role of experience and demonstrated that someone who has gained clinical experience does not automatically become an expert. For diagnostic practice and especially the education of candidate diagnosticians, it is important to know why so many clinicians do not learn from experience and why expertise is not acquired.

According to Kolodner and Simpson acquisition of expertise results from the interaction between problem solving and experience that includes feedback. In Chapter 4 it was concluded that clinicians do not learn from experience because they hardly ever receive feedback and the feedback they receive is often flawed (Garb, 1989; IJzendoorn & Bus, in press). This implies that clinicians should be trained to learn from feedback. We can agree with Garb that research should focus on helping clinicians to make the distinction between biased feedback and accurate feedback and to design training programs that meet this purpose. Garb remarks that for a task in which unbiased feedback is unavailable, clinicians should rely more on a decision-making-aid, e.g. a computer program, than on their own experience.

Furthermore, we concluded that effective interaction -and thus acquisition of expertise- only takes place if the person has access to a domain-specific knowledge base. Yet, a minimal domain-specific knowledge base is still in the making in clinical psychology. We think it would be profitable, in parallel with the accumulation of this knowledge base, to implement this basic domain knowledge in an easily accessible computerized knowledge base for clinicians in diagnostic practice. The interaction with an adequate substantial domain knowledge base of an expert system can help the clinician to learn from experience. As we mentioned before, a good classification of mental disorders, such as the empirically based system of Achenbach for the field of child psychopathology, may provide such a domain-specific knowledge base. An expert system can also help to train the clinician to learn from feedback, by keeping records of its decisions and by tracking down and explaining its failures.

It is obvious that candidate diagnosticians should also be acquainted with the domain-specific knowledge base and be trained in learning from feedback.

To conclude, if this thesis has made something clear then it is the (urgent) need for a decision aid for diagnostic practice. The publicity and commotion around some diagnostic blunders in recent cases of supposed sexual abuse once more illustrates the lack of expertise in the diagnostic field and the trickiness of diagnostics. Therefore we suggest focusing research on the design of a decision aid. A computerized diagnostic decision program or expert system appears to be beneficial as a decision aid in our view, because it can meet the different needs in the psychodiagnostic domain. It can provide the clinician with a methodologically accurate diagnostic procedure and with an adequate substantial domain knowledge base, and it can contain a case-based reasoning method in order to learn from experience. However, considering the complexity of the diagnostic task, the development of such a system may take some time. Meanwhile we think it necessary to set up training programs for diagnosticians to help them deal with the biases that trouble the diagnostic process but are inherent to being human: biased observations, biased memories, biased heuristics.
Finally, in view of these biases, diagnosticians can be considered experts in the

sense that they are so persistent in using and having confidence in their intuitive judgement. Self-confidence and self-presentation, as well as the creation and maintenance of a public image, are considered important psychological characteristics of experts that are difficult to learn and cannot easily be replaced by a computer program.

Aamodt, A. (1990). Knowledge-intensive case-based reasoning and sustained learning. In L.C. Aiello, E. Sandewall, G.Hagert, & B.Gustavsson (Eds.), *Proceedings of the European Conference on Artificial Intelligence 1990* (pp. 1-6).

Achenbach, T.M. (1966). The classification of children's psychiatric symtoms: A factor-analytic study. *Psychological Monographs, 7*, 1-37.

Achenbach, T.M. (1978a). The child behavior profile: I : Boys aged 6-11. *Journal of Consulting and Clinical Psychology, 46*, 478-488.

Achenbach, T.M. (1978b). Psychopathology of childhood: Research problems and issues. *Journal of Consulting and Clinical Psychology, 46*, 759-776.

Achenbach, T.M. (1985). *Assessment and taxonomy of child and adolescent psychopathology.* Newbury Park, CA:Sage.

Achenbach, T.M. (1988). Integrating assessment and taxonomy. In M. Rutter, A. H. Tuma, & I.S. Lann (Eds.), *Assessment and diagnosis in child psychopathology*, (pp.300-343). New York, London: The Guilford Press.

Achenbach, T.M. (1989). The derivation of taxonomic constructs: A necessary stage in the development of developmental psychopathology. In D. Cicchetti (Ed.) *Rochester Symposium on Developmental Psychopathology*, Vol.3 , Hillsdale: Erlbaum.

Achenbach, T.M., Conners, K.C., Quay, H.C., Verhulst, F.C., & Howell, C.T. (1989). Replication of empirically derived syndromes as a basis for taxonomy of child/adolescent psychopathology. *Journal of Abnormal Child Psychopathology, 17*, 299-322.

Achenbach, T.M., & Edelbrock, C.S. (1978). The classification of child psychopathology: A review and analysis of empirical efforts. *Psychological Bulletin, 85*, 1275-1301.

Achenbach, T.M., & Edelbrock, C.S. (1979). The child behavior profile: II : Boys aged 12-16 and girls aged 6-11 and 12-16. *Journal of Consulting and Clinical Psychology, 47*, 223-233.

Achenbach, T.M., & Edelbrock, C.S. (1983) *Manual for the Child Behavior Checklist and Revised Behavior Profile.* Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T.M., & McConaughy, S.H. (1987). *Empirically based assessement of child and adolescent psychopatholgy: Practical applications.* Newbury Park, CA: Sage.

Achenbach, T.M., Verhulst, F.C., Baron, G.D., & Althaus, M. (1987a). A comparison of syndromes derived from the child behavior checklist for American and Dutch Boys aged 6-11 and 12-16. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 28,* 437-445.

Achenbach, T.M., Verhulst, F.C., Baron, G.D, & Akkerhuis, G.W. (1987b). Epidemiological comparison of American and Dutch children: I.Behavior and emotional problems and competencies reported by parents for ages 4 to 16. *Journal of the American Academy of Child Psychiatry, 26,* 317-326.

Allen, M.J., & Yen, W.M. (1979). *Introduction to measurement theory.* Monterey, CA: Cole.

American Psychiatric Association (3rd rev.ed. 1987). *Diagnostic and statistical manual of mental disorders (DSM-III-R).* Washington, D.C: Author.

Blashfield, R.K. (1984). *The classification of psychopathology: Neo-Kraepelinian and quantitative approaches.* New York: Plenum Press.

Barkley, R.A. (1988). Child behavior rating scales and checklists. In M. Rutter, A.H.Tuma & I.S.Lann (Eds.), *Assessment and diagnosis in child psychopathology* (pp. 113-155). New York, London: The Guilford Press.

Breuker, J.A., Elshout, J.J., Someren, M.W. van, & Wielinga,B.J. (1986a). Hardop denken en protokolanalyse [Thinking aloud and protocol analysis]. *Tijdschrift voor Onderwijsresearch, 11,* 241-254.

Boshuizen, H.P.A. (1989). *De ontwikkeling van medische expertise; een cognitief-psychologische benadering* [On the development of medical expertise; a cognitive-psychological approach]. Unpublished doctoral dissertation, University of Maastricht, the Netherlands.

Breuker, J. & Wielinga, B. (1986). Analyse voor expertsystemen [analysis for expert systems]. In A. Nijholt & L. Steels (Eds.) *Ontwikkelingen in expertsystemen: Hulpmiddelen, technieken, methodologieën.* (pp.5.1-5.28). Den Haag: Academic Service.

Breuker, J. (Ed.) (1986). *Model-driven knowledge acquisition: Interpretation models.* Esprit Project P1098, Deliverable D1 (Task A1). Amsterdam: University of Amsterdam.

De Bruyn, E.E.J. (1985). *Psychodiagnostiek: een systematisch inleiding vanuit het klinische gezichtspunt.* [Psychodiagnostics: a systematic introduction from a clinical viewpoint]. Rotterdam: Lemniscaat.

De Bruyn, E.E.J. (1988). Besluitvorming in de klinische psychodiagnostiek [Decision making in clinical psychodiagnostics]. *Nederlands Tijdschrift voor Orthopedagogiek, 24,* 88-101.

De Bruyn, E.E.J. (1990). Clinical decision making in a multidisciplinary team. In K. Borcherding, O.I. Larichev, & D.M. Messick (Eds.), *Contemporary issues in decision making* (pp. 317-332). Amsterdam: North-Holland.

De Bruyn, E.E.J., Pijnenburg, H.M., & Van Kessel, P.G.J.(1986). Diagnostische besluitvorming in een multidisciplinair team [Diagnostic decision making in a multidisciplinary team]. *Tijdschrift voor Orthopedagogiek, 11,* 574-587.

Burton, A.M., Shadbolt, N.R., Rugg, G., & Hedgecock, A.P. (1990). The efficacy of knowledge elicitation techniques: A comparison across domains and levels of expertise. *Knowledge Acquisition, 2,* 167-178.

Bus, A.G. (1989) How are recommendations concerning reading and spelling disabilities arrived at and why do experts disagree? *Psychology in the Schools, 26,* 54-61.

Bus, A.G., & Kruizinga, T.H. (1989). Diagnostic problem-solving behavior of expert practitioners in the field of learning disabilities. *Journal of School Psychology, 27,* 277-287.

Cantor, N., Smith, E.E., French, R.deS, & Mezzich, J. (1980). Psychiatric diagnosis as prototype categorization. *Journal of Abnormal Psychology, 89,* 181-193.

Carroll, B. (1987). Expert systems for clinical diagnosis: Are they worth the effort? *Behavioral Science, 32,* 274-292.

Chan, D.W., & Jackson, D.N. (1982) Individual differences in the perception and judgement of psychopathology. *Multivariate Behavioral Research, 17,* 3-32.

Chapman, L.J., & Chapman, J. (1982). Test results are what you think they are. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases*(pp. 239-248). Cambridge: Cambridge University Press.

Clancey, W.J. (1985). Heuristic classication. *Artificial Intelligence, 27,* 289-350.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37-46.

Costello, E.J. (1982, October). *Clinical decision making in child psychiatry.* Paper presented at the 29th meeting of the American Academy of Child Psychiatry, Washington, DC.

Dawes, R.M., Faust, D., & Meehl, P.E. (1989). Clinical versus actuarial judgement. *Science, 243,* 1668-1673.

Dowie, J. A., & Elstein, A.S. (Eds.).(1988). *Professional judgement: A reader in clinical decision making.* Cambridge: University Press.

Drenth, P.J.D. (1975). *Inleiding in de testtheorie* [Introduction to test theory]. Deventer: Van Loghum Slaterus.

Dreyfus, H.L., & Dreyfus, S.E. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer.* New York: The Free Press.

DSM-IV Work group (1990). Child and Adolescent Psychiatric Disorders in DSM-IV: Issues facing the work group [Brief communication]. *Journal of the American Academy of Child and Adolescent Psychiatry,* p.830.

Elio, R., & Scharf, P.B. (1990). Modelling novice-to-expert shifts in problem solving strategy and knowledge organisation. *Cognitive Science, 14,* 579-640.

Elstein, A.S., Shulman, L.S., & Sprafka, S.A. (1978) *Medical problem solving: An analysis of cinical reasoning.* Cambridge: Harvard University Press.

Elstein, A.S., & Bordage, G. (1988). Psychology of clinical reasoning. In J.A. Dowie & A.S. Elstein (Eds.), *Professional judgement: a reader in clinical decision making* (pp. 109-129).Cambridge: Cambridge University Press.

Edelbrock, C., & Costello, A.J. (1988). Structured psychiatric interviews for children. In M. Rutter, A.H. Tuma, & I.S. Lann (Eds.), *Assessment and diagnosis in child psychopathology* (pp. 87-112). New York, London: The Guilford Press.

Ericsson, K.A., & Simon, H.A. (1984).*Protocol analysis: Verbal reports as data.* Cambridge: MIT Press.

Eysenck, H.J. (1960). Classification and the problem of diagnosis. In H.J. Eysenck (Ed.), *Handbook of abnormal psychology* (pp. 1-31).London: Pitman Press.

Farrell, A. D. (1984). When is a computerized assessment system ready for distribution? Some standards for evaluation. In M. D. Schwartz (Ed.), *Using computers in clinical practice* (pp.185-189). New York: Hayworth Press.

Feinstein, A.V. (1977). A critical overview of diagnosis in psychiatry. In V.H. Rakoff, H.C. Stancer, & H.B. Kedward (Eds.), *Psychiatric Diagnosis* (pp.189-228) London: Macmillan.

Garb, H.W.(1989). Clinical judgement, clinical training, and professional experience. *Psychological Bulletin, 105,* 387-396.

Geraats, E., & Vermeulen, F. (1991). Onderzoek naar de discrepantie tussen een empirische en een klinische classificatie [Study on the discrepancy between an empirical and a clinical classification].Unpublished master's thesis, University of Nijmegen, The Netherlands.

De Groot, A.D. (1961). *Methodologie: Grondslagen van onderzoek en denken in de gedragswetenschappen* [Methodology: Foundations of reseach and cognition in behavioral science]. 's-Gravenhage: Mouton.

De Groot, A.D. (1965). *Thought and choice in chess.* The Hague: Mouton.

Hamm, R.M. (1988). Clinical intuition and clinical analysis: Expertise and the cognitive continuum. In J.A. Dowie & A.S. Elstein (Eds.), *Professional judgement: a reader in clinical decision making* (pp. 78-105). Cambridge: Cambridge University Press.

Horowitz, L.M., Wright, J.C., Lowenstein, E., & Parad, H.W. (1981a). The prototype as a construct in abnormal psychology: 1. Method for deriving prototypes. *Journal of Abnormal Psychology, 90,* 568-574.

Horowitz, L.M., Post, D.L., French, R. de S., Allis, K.P., & Siegelman, E.Y. (1981b). The prototype as a construct in abnormal psychology: 2. Clarifying disagreement in psychiatric judgements. *Journal of Abnormal Psychology, 90,* 575-585.

Jackson, P. (1987). *Expertsystemen* [Expert systems]. Amsterdam: Addison-Wesley.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.)(1982). *Judgement under uncertainty: Heuristics and biases.* Cambridge: Cambridge University Press.

Kendell, R.E. (1973). Psychiatric diagnoses: A study of how they are made. *British Journal of Psychiatry, 122,* 437-445.

Kleinmuntz, B. (1968). The processing of clinical information by man and machine. In B. Kleinmuntz (Ed.), *Formal representation of human judgement* (pp. 149-186). New York: John Wiley & Sons, Inc.

Kleinmuntz, B. (1990). Why we still use our heads instead of formulas: Toward an integrative approach. *Psychological Bulletin, 107,* 296-310.

Kolodner, J.L. (1984). Towards an understanding of the role of experience in the evolution from novice to expert. In M.J. Coombs (Ed.), *Developments in expert systems: From a special issue of the International Journal of Man-Machine Studies* (pp. 95-116). London: Academic Press.

Kolodner, J.L., & Simpson, R.L. (1986). Problem solving and dynamic memory. In J.L. Kolodner & C.K.Riesbeck (Eds.), *Experience, memory and reasoning* (pp. 99-114). Hillsdale, NJ: Erlbaum.

Koster van Groos, G.A.S. (1989). *Beknopte handleiding bij de diagnostische criteria van de DSM-III- R* [Quick Reference to the Diagnostic Criteria from DSM-III-R], Amsterdam: Lisse, Swets & Zeitlinger.

Kooij, R. van der, & Knijff, H.K. (Eds.) (1986) *Tussen klacht en behandeling,* diagnostiseren in de othopedagogiek.[Between complaint and treatment; diagnosing in special education] Nijmegen: Dekker & Van de Vegt.

Koot, J.M., & Verhulst, F.C., (1990). Stabiliteit van door ouders gerapporteerd probleemgedrag bij 4 tot 12 jarigen [Stability of problem behavior of children aged 4 - 12 reported by parents]. *Kind en Adolescent, 11,* 5-18.

Krol, N.P.C.M. (1990). *Handleiding codering hardop-denkprotocollen* Coding manual]. Internal report, University of Nijmegen, The Netherlands: N.I.C.I., Department of Special Education.

Krol, N.P.C.M., & De Bruyn, E.E.J. (1990a). De inhoudsvaliditeit van de Child Behavior Checklist [The content validity of the CBCL]. *Nederlands Tijdschrift voor Opvoeding, Vorming en Onderwijs, 6*, 322-332.

Krol, N.P.C.M., & De Bruyn, E.E.J. (1990b, November). Validity of the CBCL [Letter to the editor]. *Journal of the American Academy of Child and Adolescent Psychiatry*, p.986.

Kruizenga, T.H, & Bus, A.G. (1990). Expert and novice performance at diagnosing children with reading and spelling problems. University of Groningen, The Netherlands: Department of Education.

Kuipers, B., & Kassirer, J.P. (1984). Causal reasoning in medicine: Analysis of a protocol. *Cognitive Science, 8*, 363-385.

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 44*, 159-174.

McCauley, R.N. (1987) The role of theories in a theory of concepts. In U. Neisser (Ed.) *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp.288-309). Cambridge: Cambridge University Press.

Mechelen, Van I., & Boeck, de P. (1989). Implicit taxonomy in psychiatric diagnosis: A case study. *Journal of Social and Clinical Psychology, 8*, 276-287.

Meehl, P.E. (1973). *Psychodiagnosis: Selected Papers.* Minneapolis: University of Minnesota Press.

Neisser, U. (Ed.)(1987). *Concepts and conceptual development: Ecological and intellectual factors in categorization.* Cambridge: University Press.

Newell, A., & Simon, H.A. (1972). *Human problem solving.* Englewood Cliffs, N.J.: Prentice-Hall.

Nurcombe, B., & Fitzhenry-Coor, I.(1986). A review of research into clinical reasoning. In Nurcombe, B. & Gallagher, R.M., (Eds.), *The clinical process in Psychiatry: Diagnosis and management plannning* (pp. 605-631). Cambridge: University Press.

Patil, R.M. (1987). A case study on evolution of system builing expertise: Medical diagnosis. In W.E.L. Grimson & R.S. Patil (Eds.), *AI in the 1980's and beyond* (pp. 75-108). Cambridge: MIT Press.

Popping, R. (1983). *Overeenstemmingsmaten voor nominale data.* [Measures of agreement for nominal data]. Doctoral dissertation, University of Groningen, The Netherlands. Meppel: Krips Repro.

Pijl, Y.L. (1989). *Het toelatingsonderzoek in het LOM- en MLK-onderwijs* [Admittance research in special education]. Doctoral dissertation, University of Groningen, The Netherlands: RION.

Rispens, J. (1986). Diagnostiek en classificatie. [Diagnostics and classification]. In L. Levering & J. Van Weelden (Eds.), *Over het bijzondere van de orthopedagogiek* (pp. 15-44). Groningen, The Netherlands: Wolters Noordhoff.

Rispens, J., Carlier, E., & Schoorl, P.(Eds.) (1984) Diagnostiek in de hulpverlening: methodologische aspecten van de diagnostiek van opvoedings- en leerproblemen. [Diagnostics in mental health services: Methodological aspects of the diagnosis of rearing- and learning problems] Lisse: Swets & Zeitlinger.

Rosch, E., Mervis, C.B., Wayne, D.G., Johnson, D.M., & Boyes-Bream, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382-439.

Ruijssenaars, A.J.J.M., & De Bruyn, E.E.J. (1987). Expertsystemen in de klinische diagnostiek [Expert systems in clinical diagnostics]. *De Psycholoog, 22*(1), 6-9.

Rutter, M. (1975). *Helping troubled children.* New York, London: Plenum Press.

Rutter, M., & Gould, M. (1985). Classification. In M.Rutter, & L.Hersov, (Eds.), *Child and adolescent psychiatry* (pp. 304-321). London: Blackwell.

Rutter, M., Tuma, A.H., & Lann, I.S. (Eds.) (1988).*Assessment and diagnosis in child psychopathology.* New York, London: The Guilford Press.

Rutter, M., & Tuma, A.H. (1988). Diagnosis and classification: Some outstanding issues. In M. Rutter, A.H. Tuma, & I.S.Lann (Eds.), *Assessment and diagnosis in child psychopathology* (pp.437-464). New York, London: The Guilford Press.

Schwartz, M. D. (1984). *Using computers in clinical practice.* New York: Hayworth Press.

Schwartz, S., Griffin, T., & Fox, J. (1989). Clinical expert systems versus linear models: Do we really have to choose? *Behavioral Science, 34,* 305-311.

Shanteau, J. (1988). Psychological characteristics and strategies of expert decision makers. *Acta Psychologica, 68,* 203-215.

Shanteau, J. (in press). Psychological characteristics and strategies of experts. In G.Wright & F. Bolger (Eds.), *Expertise and decision support.* New York: Plenum.

Shaw, M.L.G., & Woodward, J.B. (1990). Modelling expert knowledge. *Knowledge Acquisition, 2,* 179-206.

Slotboom, A. (1987). *Statistiek in woorden, de meest voorkomende termen en technieken* [Statistics in words: The most frequent terms and techniques]. Groningen: Wolters-Noordhoff.

Svensson, O. (1989). Eliciting and analyzing verbal protocols in process studies of judgement and decision making. In H. Montgomery & O. Svenson (Eds.), *Process and structure in human decision making.* Chicester: John Wiley & Sons.

Ten Cate, I., & Hoge, T. (1989) Child Behavior Checklist: *Als diagnostisch instrument om gedragsproblemen te inventariseren bij kinderen met motorische stoornissen* [The CBCL as a diagnostic instrument to collect behavioral problems in children with motor dysfunctions.] Unpublished master's thesis, University of Nijmegen, The Netherlands.

Traudes, M.A. (1987, august). *Project "Eerstelijns expertsystemen in de diagnostiek"* [Expert systems in diagnostics: Report no.1]. Internal report, University of Nijmegen, Department of Special Education. The Netherlands.

Turk, D.C., & Salovey, P. (Eds.) (1988). *Reasoning, inference and judgement in clinical psychology.* London: The Free Press.

Verhulst, F.C. (1985). *Mental health in Dutch children.* Meppel: Krips.

Verhulst, F.C., Achenbach, T.M., Althaus, M., & Akkerhuis, G.W. (1988). A comparison of syndromes derived from the Child Behavior Checklist for American and Dutch girls aged 6-11 and 12-16. *Journal of Child Psychology and Psychiatry, 29,* 879-895.

Verhulst, F.C., Koot, J.M., Akkerhuis, G.W., & Veerman, J.W.(1990). *Child Behavior Checklist* (CBCL). Assen: Van Gorcum.

Vermande, M. M. (1991) Correspondence between DSM-III-R disorders and CBCL synodromes as a basis for case construction. Unpublished manuscript, Department of Special Education, University of Nijmegen, The Netherlands.

Westmayer, H. (1975). The diagnostic process as a statistical-causal analysis. *Theory and Decision, 6,* 57-86.

Wiggins, J.S. (1971). *The psychology of personality.* Reading, Mass: Addison-Wesley.

Woody, R.H. (Ed.) (1980). *Encyclopedia of clinical assessment.* San Francisco: Jossey-Bass.

Woolf, H.B. (Ed.) (1977). *Webster's New Collegiate Dictionary.* Springfield, MA.: Merriam.

Wouters, F. (1989). Van vazen naar verklaringen: een trainings- en ondersteuningssysteem voor Bayesiaanse hypothese-evaluatie in de psychodiagnostiek. [From vases to explanations: a training and support system for Bayesian hypothesis-evaluation in psychodiagnostics]. Unpublished master's thesis, University of Nijmegen, The Netherlands.

IJzendoorn, Van W.J.E., & Bus, A.G. (in press). How valid are experts' prognoses on children with learning problems? *Journal of School Psychology.*

The referral of a child to a diagnostician is the start of a complex information-processing and problem-solving process during which decisions are continually being made. Research on clinical decision making has repeatedly shown the fallibility of this decision process. This is not surprising in view of the large amount of information involved and the complexity of the task. To improve diagnostic practice it is necessary to develop (computerized) diagnostic decision support systems. Especially the expert system as a decision aid is thought to offer promising possibilities for improving the quality of current psychodiagnostic practice (Ruyssenaars & De Bruyn, 1987).

This thesis is part of a research project that aims to develop expert systems for the domain of psychodiagnostics, and is concerned with the search for expert knowledge. In particular one important aspect of psychodiagnostics, the classification of problem behaviors, is addressed.

The knowledge an expert system includes can be roughly divided into two types: public and private knowledge. Public knowledge includes the types of information found in textbooks or handbooks: definitions, facts and theories. Private knowledge consists of rules of thumb or heuristics that enable experts to make educated guesses when necessary. This thesis addresses the public knowledge (Chapters 2 and 3), as well as the private knowledge (Chapters 4 and 5) involved in the task of classification of problem behaviors reported in children in the age of 6-11 years.

Chapter 1 specifies the notion of an expert, reviews research on clinical reasoning, focuses on expert-novice differences and the role of experience. An expert is defined as someone (a) who has finished his or her professional education and is familiar with the public knowledge in the domain and (b) who has gained clinical experience and from this experience has learned how to apply and use that knowledge. Research shows that cognitive biases influence the reasoning processes of expert clinicians. Therefore, the approach of modeling the clinical expert, an approach often used in expert system development, requires caution. The exploration of private knowledge in the domain of psycho-diagnostics presupposes a validation of that knowledge. However, it is difficult to define objective criteria to validate the expert knowledge in the domain of psychodiagnostics (Shanteau, 1988). Research that has focused on expert-novice differences in order to determine the value of expert knowledge, offers an inconsistent picture of the relationship between experience, problem solving, and expertise. This raises questions on the role of experience in the expert's reasoning processes. According to Kolodner and Simpson (1986), the first role of experience is to refine and modify the reasoning process. Successful experiences reinforce already known rules or previous hypotheses, whereas failures require analysis of the reasoning and knowledge that was used and modification of faulty rules and knowledge. The second role of experience is to provide exemplars upon which to base later decisions. Analogies to previous cases guide and focus later decision making. However, the added value Kolodner and many others assume experience

provides, is not always demonstrated in studies (Garb, 1989). In this thesis the role of experience for a psychodiagnostic classification task is examined in Chapter 4 and Chapter 5.

Chapter 2 presents the public knowledge that is available in the domain of psychodiagnostic classification. Psychodiagnostic classification is defined as the ordering or grouping of problematic behaviors (the phenomena) of the individual case into syndromes (the groups or types). It is part of the diagnostic process and takes place during the second phase of the diagnostic cycle described by De Bruyn, i.e. the problem identification (1985, 1990). Although classification as a cognitive process has received little attention in research on psychodiagnostics, several classification systems have been developed to assist the diagnostician during this task. Some basic requirements for an adequate classification system for child psychopathology, were proposed by Rutter and Gould (1985). Two systems that seem to be most important for the domain of child psychopathology, i.e. the DSM-III-R system and the CBCL system, are described and compared with respect to some general issues, i.e. (a) categorical versus dimensional, (b) categories versus prototypes, (c) reliability, and (d) validity. In view of our ultimate goal, the construction of a diagnostic knowledge based system, it is concluded that the empirically based assessment approach of the CBCL system is the most solid one to be used as a database for a such a system.

Chapter 3 reports an empirical study into the content validity of the CBCL. Despite the favourable judgement of the empirically based assessment method, it is questioned whether the CBCL is sufficiently representative to be used as a data base in a computerized diagnostic knowledge system. If certain problem behaviors are not represented in the CBCL they will not be detected in a forthcoming case. The study demonstrates that the items of the CBCL do not cover the broad range of problem behaviors reported in diagnostic reports. It is concluded that several items should be added. These findings subsequently raise the question of how to decide whether the added items or symptoms are part of the empirically derived syndromes and illustrate a functional problem in the use of classification systems: how to deal with individual problem behavior that doesn't fit into the classes of a system. Following Kleinmuntz, who was able to extract the decision rules an expert was using in the classification of MMPI-profiles, we focused our research on the classification process of experienced diagnosticians in order to obtain practical guidelines for dealing with such a classification problem. Chapter 4 reports the study on how experienced diagnosticians, referred to as experts, classify individual problem behavior, in comparison with novices. Inspired by Kolodner's model of the evolution of expertise (described in Chapter 1) we examined the differences between experts and novices. We identified seven cognitive operations that experienced diagnosticians use in classifying problem behavior: (a) asking or giving information, (b) associating, (c) abstracting or labelling, (d) explaining, (e) neutral matching, (f) identifying, and (g) differentiating. Results show that experienced diagnosticians and novices differed

in the use of five of these cognitive operations. This finding suggests that experience has taught the diagnostician to change his reasoning process by questioning, abstracting or labelling the problem behavior more often during classification and supports the change in the reasoning process that Kolodner's model postulates. But we also found that the outcomes of the classification and the typicality ratings given by the experts are not different from those of the novices. This last finding disagrees with the change in domain knowledge Kolodner's model hypothesizes and casts doubt on the expert status of experienced diagnosticians. Why did persons who received a lot of training and do have a lot of experience (some over 20 years) in the clinical psychological domain, not learn enough to become full-fledged experts in the sense of Kolodner's model? We concluded that an effective interaction between problem solving on the one hand and experience (including feedback) on the other hand can only take place if a person has access to background knowledge in a domain-specific knowledge base. However, in the clinical psychological domain the minimal domain-specific knowledge base that candidate experts should use in order to learn from experience is still in the making.

The results of this study show that experts and novices have the same conceptions about the co-occurrence of problem behavior, which indicates that the prototypes of experts and novices are much alike. We subsequently focused our research on an evaluation of intuitive prototypes, i.e., the classification knowledge as represented in cognitive structures of experienced diagnosticians. In Chapter 5 the intuitive classification structures of experienced diagnosticians are examined, in particular the empirical content of their intuitive prototypes. We compared intuitive prototypes of diagnosticians with empirical prototypes derived by multivariate analysis. Results show that the diagnostician's intuitive clusters correspond very poorly with the empirical syndromes. The similarity between behaviors that clinicians judge as co-occurring with an incompletely specified syndrome and behaviors that empirically co-occur with that same syndrome was likewise low. We conclude that intuitive prototypes do not correspond with empirical prototypes, in other words, that the empirical content of intuitive prototypes is minimal. The effect of expertise on intuitive classification knowledge should be examined in more detail in the future. As yet we do not know what sort of unevenness or bias expertise creates in a taxonomy as used, e.g., by a specialist in depression; or on the classification judgement by that specialist. Therefore clinicians and candidate clinicians are advised to use the available empirical classification knowledge on the co-occurrence of problem behavior and not to rely on their own possibly biased memory.

In the final chapter the main findings of this thesis and its implications for the notion of an expert and expert knowledge are discussed, in particular in relation to the development of expert systems. Also some suggestions for future research and for the diagnostic practice are made.

This thesis used two different criteria to evaluate the private classification

knowledge of experienced diagnosticians, i.e. the experience criterion, (expert-novice differences), and an empirical criterion, (the empirical classification knowledge as provided by the CBCL). The results of these studies cast doubt on the expert status of diagnosticians. Instead of concluding that there are no experts in the psychodiagnostic domain, it is suggested to expand the notion of an expert with a new aspect, i.e. the methodological correctness of his or her diagnostic procedure. At this moment, the lack of objective standards for validating the outcome of the diagnostic process and the lack of scientific knowledge in child psychopathology seems to demand an explicit or transparent diagnostic procedure (for instance the normative diagnostic cycle model as described in Chapter 2) to evaluate the expertise of the clinician. It is assumed that the experienced diagnosticians will gain from the application of an accurate diagnostic procedure to real cases, which will (it may be hoped) result in expertise.

Results of this and other studies (Garb, 1989) implicate that it is not profitable to base an expert system on the knowledge of experienced diagnosticians. If we want a computerized aid or system to be considered an expert system, its content should meet the same criteria that we applied to the notion of an expert. This means that the system should contain (a) a methodologically correct diagnostic procedure, for instance the normative diagnostic cycle, (b) adequate domain knowledge, for instance the CBCL system, and (c) a reasoning method that learns from experience, for instance case-based reasoning.

As to future research it is suggested to select diagnosticians more carefully in studies that aim to find expertise and to further investigate the effect of the elicitation method.

For the diagnostic practice it is considered important to stimulate an effective interaction between problem solving and experience, and thus acquisition of expertise. Therefore clinicians and candidate clinicians should be trained to learn from feedback, and basic domain knowledge should be implemented in an easily accessible computerized knowledge base.

De verwijzing van een kind naar een diagnost is de start van een complex informatieverwerkings- en besluitvormingsproces. Onderzoek naar klinische besluitvorming heeft herhaaldelijk de onvolmaaktheid van dit besluitvormingsproces aangetoond. Dat is niet verwonderlijk gezien de grote hoeveelheid informatie die bij dit proces betrokken is en de complexiteit van de taak. De ontwikkeling en het gebruik van (gecomputeriseerde) diagnostische beslissingsondersteunende systemen is nodig om de diagnostische praktijk te verbeteren (Ruyssenaars & De Bruyn, 1987).

Dit proefschrift maakt deel uit van een onderzoeksproject dat als doel heeft expertsystemen ten behoeve van de psychodiagnostiek te ontwikkelen. Het proefschrift richt zich vooral op een belangrijk onderdeel van het psychodiagnostische proces, namelijk de classificatie van probleemgedragingen.

De kennis in een expertsysteem kan ruwweg worden verdeeld in twee typen: publieke kennis en private (niet publieke) kennis. Publieke kennis omvat informatie zoals die staat in studie- of handboeken: definities, feiten en theorieën. Private kennis bestaat uit de vuistregels of heuristieken die een expert gebruikt om -indien nodig- een gefundeerde schatting te maken. Dit proefschrift richt zich zowel op de publieke kennis (hoofdstukken 2 en 3), als op de private kennis (hoofdstukken 4 en 5) voor de taak van classificatie van probleemgedragingen van 6-11 jarige kinderen.

In hoofdstuk 1 wordt het begrip expert gespecificeerd en wordt onderzoek naar klinisch redeneren besproken, in het bijzonder met de nadruk op expert-noviet verschillen en de rol van ervaring. Een expert is gedefinieërd als iemand (a) die zijn of haar professionele opleiding beëindigd heeft en bekend is met de publieke kennis van het domein en (b) die klinische ervaring heeft opgedaan en daarvan heeft geleerd hoe die publieke kennis toe te passen en te gebruiken. Onderzoek heeft aangetoond dat cognitieve vertekeningen (biases) het redeneerproces van de diagnost beïnvloeden. Daarom moet het modelleren van de klinische expert -een benadering die vaak gebruikt wordt bij de ontwikkeling van een expertsysteem-, met de nodige voorzichtigheid geschieden. Een inventarisatie van de private kennis in het psychodiagnostische domein vereist tevens een validiatie van die kennis. Het is echter moeilijk om objective criteria te definiëren waarmee de expertkennis in het psychodiagnostische domein gevalideerd kan worden (Shanteau, 1988). Onderzoek waarin expert-noviet verschillen gebruikt werden als evaluatiecriterium voor expertkennis, laat geen eenduidige relatie zien tussen ervaring, probleemoplossend gedrag en expertise. Dit roept vragen op over de rol van ervaring in het redeneerproces van de expert. Volgens Kolodner en Simpson (1986) verfijnt en modificeert ervaring het redeneerproces. Succesvolle ervaringen bekrachtigen de reeds bekende regels of hypothesen. Fouten daarentegen vereisen een analyse van het redeneerproces en de kennis die daarbij gebruikt werd, en een verandering van de foute regels en kennis. Als tweede rol van ervaring noemen zij het verschaffen van voorbeelden waarop latere beslissingen gebaseerd kunnen worden. Overeenkomsten met vroegere casus begeleiden en

richten latere besluitvorming. De toegevoegde waarde die ervaring volgens Kolodner en vele anderen verschaft, wordt lang niet altijd aangetoond in empirische studies (Garb, 1989). In dit proefschrift is de invloed van ervaring op de uitvoering van een psychodiagnostische classificatie taak onderzocht, en beschreven in hoofstuk 4 en 5.

In hoofdstuk 2 hebben we ons gericht op de beschikbare publieke kennis op het gebied van psychodiagnostische classificatie. Psychodiagnostische classificatie is gedefinieërd als het ordenen of groeperen van problematische gedragingen (de symptomen) van de individuele casus (een kind) in syndromen (de groepen of typen). Classificatie maakt deel uit van het diagnostische proces en vindt plaats gedurende de 'probleem identificatie', de tweede fase van de diagnostische cyclus volgens De Bruyn (1985, 1990).

Classificatie als cognitieve activiteit heeft weinig aandacht gekregen in onderzoek naar psychodiagnostiek, maar er zijn wel verschillende classificatiesystemen ontwikkeld om de diagnost bij deze taak te assisteren. Rutter en Gould (1985) hebben beschreven aan welke eisen een geschikt classificatiesysteem moet voldoen. Wij hebben twee systemen beschreven die belangrijk zijn voor de psychopathologie bij kinderen: het DSM-III-R systeem en het CBCL systeem. Deze twee systemen zijn vergeleken op een viertal aspecten: (a) categorisch versus dimensioneel, (b) categorieën versus prototypen, (c) betrouwbaarheid, en (d) validiteit. De empirisch gefundeerde classificatiekennis van het CBCL systeem blijkt de meest solide basis te zijn voor een diagnostisch kennissysteem.

In hoofdstuk 3 wordt een empirische studie naar de inhoudsvaliditeit van de CBCL beschreven. We onderzochten of het itembestand van de CBCL voldoende representatief is om gebruikt te worden als databestand in een gecomputeriseerd diagnostisch kennissysteem. Als bepaalde probleemgedragingen niet voorkomen in het systeem kunnen ze ook niet bij een nieuwe casus ontdekt worden. De studie toont aan dat de items van de CBCL de brede range van probleemgedragingen niet volledig dekken. Om de CBCL en de bijbehorende empirische kennis te kunnen gebruiken als een vaste component van een diagnostisch kennissysteem is een uitbreiding van het itembestand gewenst. De vraag is dan of de toegevoegde, extra items al dan niet passen bij de empirisch verkregen syndromen. Deze vraag illustreert een probleem dat inherent is aan het gebruik van classificatiesystemen: wat moet je doen met individueel probleemgedrag dat niet in de categorieën van het systeem past. Om praktische richtlijnen te verkrijgen voor het omgaan met een dergelijk classificatieprobleem hebben we ons onderzoek gericht op het classificatieproces van ervaren diagnosten. Dit is gedaan in navolging van Kleinmuntz, die in staat was om de beslisregels te achterhalen die een expert gebruikte bij het classificeren van MMPI-profielen.

De studie beschreven in hoofdstuk 4, onderzoekt hoe ervaren diagnosten individueel probleemgedrag classificeren, in vergelijking met novieten. Geïnspireerd door het model van Kolodner over de evolutie van expertise (beschreven in hoofdstuk 1) is het verschil in classificatieproces en-uitkomst tussen experts en

novieten onderzocht. We identificeerden zeven cognitieve operaties die ervaren diagnosten gebruikten tijden het classificeren van probleemgedragingen: (a) vragen naar of geven van informatie, (b) associëren, (c) abstraheren of een label geven, (d) verklaren, (e) neutraal vergelijken, (f) identificeren, en (g) differentiëren. De resultaten laten zien dat ervaren diagnosten en novieten verschillen in het gebruik van 5 cognitieve operaties. Deze bevinding suggereert dat de diagnost door ervaring geleerd heeft om zijn redeneerproces te veranderen door het probleemgedrag meer te bevragen, meer te abstraheren en van een label te voorzien tijdens het classificeren. Dit ondersteunt de veronderstelling van Kolodner dat ervaring het redeneerproces verandert. Maar we vonden ook dat de opbrengst van de classificatie en de 'typicaliteits beoordeling' door experts niet verschilt van die door de novieten. Dit laatste resultaat is niet in overeenstemming met de verandering in domeinkennis die het model van Kolodner veronderstelt en hiermee wordt de expertstatus van ervaren diagnosten in twijfel getrokken. Waarom hebben personen die een opleiding hebben gehad en veel ervaring hebben in de klinische praktijk (sommige meer dan 20 jaar) niet voldoende geleerd om volledige experts te worden in de betekenis die Kolodner op basis van haar ervaringsmodel aan de term expert geeft? Wij komen tot de conclusie dat er geen effectieve interactie tussen probleemoplossen en ervaring heeft plaats gehad omdat de diagnosten niet genoeg feedback kregen over hun genomen beslissingen en omdat ze niet konden beschikken over ondersteunende kennis in een domeinspecifiek kennisbestand. We veronderstellen dat voor de klinische psychologie een minimaal domeinspecifiek kennisbestand nog in ontwikkeling is.

De resultaten van deze studie tonen ondermeer aan dat experts en novieten dezelfde concepten hebben over het samen-voorkomen van probleemgedragingen. Dit geeft aan dat de intuitieve prototypen van experts en novieten veelal hetzelfde zijn.

In hoofdstuk 5 hebben we de intuitieve prototypen van ervaren diagnosten onderzocht, in het bijzonder de empirische dekking ervan. We hebben intuitieve prototypen vergeleken met – via multivariate analyse verkregen – empirische prototypen. We lieten diagnosten drie classificatietaken uitvoeren en vergeleken de opbrengst met de empirische classificatiekennis behorende bij de CBCL. Resultaten laten zien dat de intuitieve clusters van de diagnost weinig overeenkomsten vertonen met de empirische clusters. We concluderen dat de intuitieve prototypen niet overeenkomen met empirische prototypen of anders gezegd dat de empirische dekking van de intuitieve prototypen minimaal is. Aan diagnosten wordt geadviseerd om gebruik te maken van de beschikbare empirische kennis over het samenvoorkomen van probleemgedragingen en niet teveel te vertrouwen op hun eigen, mogelijke vertekende (biased) geheugen.

In het laatste hoofdstuk zijn de belangrijkste bevindingen van het onderzoek en de implicaties voor de begrippen van een expert en expertkennis besproken, speciaal in het licht van de ontwikkeling van een expertsysteem. Ook zijn er

aanbevelingen gedaan voor verder onderzoek en voor de diagnostische praktijk.

Dit proefschrift heeft twee criteria gebruikt om de private, classificatiekennis van ervaren diagnosten te evalueren, namelijk het criterium ervaring (expert-noviet verschil) en een empirisch criterium (de empirische kennis van de CBCL). Door de resultaten van de studies is de expertstatus van ervaren diagnosten in twijfel getrokken. In plaats van te concluderen dat er geen experts zijn op het gebied van de psychodiagnostiek, doen we de suggestie om het begrip expert uit te breiden met een nieuw aspect, namelijk de methodologische correctheid van zijn of haar diagnostische procedure. Het gemis aan objectieve criteria waarmee de opbrengst van het diagnostische proces gevalideerd kan worden en het gebrek aan wetenschappelijke kennis op het gebied van psychopathologie bij kinderen, maken het belangrijk om de expliciteit en doorzichtigheid van de toegepaste diagnostische procedure te betrekken bij de beoordeling van de expertise van een diagnost. We veronderstellen dat de diagnost profijt zal hebben van het gebruik van een correcte diagnostische procedure op echte casussen en dat dit dan hopelijk zal resulteren in expertise.

De resultaten van deze en andere studies (Garb, 1989) impliceren dat het niet zinvol is een expertsysteem te baseren op de private kennis van diagnosten. Als we een gecomputeriseerd diagnostich kennissysteem willen beschouwen als een expertsysteem dan moet de inhoud van dat systeem voldoen aan dezelfde eisen die we aan het begrip expert hebben gesteld. Dit houdt in dat het systeem (a) over een methodologische correcte diagnostische procedure moet beschikken, bijvoorbeeld de normatieve diagnostische cyclus, (b) geschikte domeinkennis moet bevatten, en (c) over een redeneermethode moet beschikken waarmee het kan leren van ervaring, bijvoorbeeld 'case-based reasoning'.

Voor verder onderzoek wat als doel heeft om expertise te vinden op het gebied van de klinische psychologie bevelen we aan de diagnosten zorgvuldig te selecteren en het effect van de uitlokkingsmethode op de verkregen kennis nader te onderzoeken.

Voor de diagnostische praktijk is het belangrijk om een effectieve interactie te stimuleren tussen probleemoplossen en ervaring, en zo het verkrijgen van expertise te bevorderen. Hiervoor moeten diagnosten getraind worden om optimaal gebruik te maken van feedback en moet basale domeinkennis geïmplementeerd worden in een gecomputeriseerd kennisbestand dat makkelijk toegankelijk en te gebruiken is.

Because the study was performed in Dutch, some details and data of the study are given in Dutch in the Appendices.

Appendix A

The empirical syndromes of the CBCL as found in a Dutch clinical population*.

Syndromes of girls aged 6-11

1.	**AGGRESSIVE (AGRESSIEF)**	
95.	Driftbuien of snel driftig	.71
3.	Spreekt veel tegen of maakt veel ruzie	.70
86.	Koppig, stuurs of prikkelbaar	.70
22.	Is thuis ongehoorzaam	.68
104.	Is erg luidruchtig	.66
87.	Verandert plotseling van stemming	.65
88.	Mokken, pruilen	.64
68.	Schreeuwt veel tegen of gilt veel	64
19.	Eist veel aandacht op	.61
93.	Praat te veel	.58
27.	Snel jaloers	.57
94.	Plaagt veel	.54
10.	Kan niet stil zitten, onrustig of overactief	.51
26.	Lijkt zich niet schuldig te voelen na zich misdragen te hebben	.51
109.	Dreinerig, jengelig	.48
74.	Speelt de clown, doet raar of 'gek' om de aandacht te trekken	.46
41.	Impulsief of onnadenkend gedrag	.44
90.	Vloeken, schuttingtaal	.43
16.	Wreed, pesterig of gemeen voor anderen	.43
45.	Zenuwachtig, gespannen	.42
8.	Kan zich niet concentreren, kan niet lang de aandacht bij iets houden	.40
2.	**CRUEL (WREED)**	
81.	Steelt van huis	.60
82.	Steelt buitenshuis	59
43.	Liegen of bedriegen	.57
21.	Vernielt spullen van andere gezinsleden of andere kinderen	.54
16	Wreed, pesterig of gemeen voor anderen	.48
37.	Vecht veel	.44
48.	Andere kinderen mogen haar niet	.43
94.	Plaagt veel	.40
25.	Kan niet goed opschieten met andere kinderen	.39
57.	Valt andere aan of valt anderen lastig	39
26.	Lijkt zich niet schuldig te voelen na zich misdragen te hebben	.39
20.	Vernielt eigen spullen	.39
53.	Eet te veel	.38
7.	Opscheppen, stoer doen	.37
3.	**DEPRESSIVE (DEPRESSIEF)**	
35.	Voelt zich waardeloos of minderwaardig	.67
103.	Ongelukkig, verdrietig of gedeprimeerd	.64
12.	Klaagt over zich eenzaam voelen of in de steek gelaten voelen	.63
33.	Klaagt erover of heeft het geval dat niemand van haar houdt	.60
52.	Te veel last van schuldgevoel	.57
112.	Maakt zich zorgen	.56
32.	Vindt dat zij perfect moet zijn	.52
34.	Heeft het gevoel dat anderen het op haar gemunt hebben	.51
31.	Is bang dat zij iets ondeugends of slechts zou kunnen denken of doen	.47

** The items defining the syndromes are presented in descending order of factor loading.*

71.	Schaamt of geneert zich snel	.45
30.	Is bang om naar school te gaan	.40
38.	Wordt veel geplaagd	.36
45.	Is te angstig of te bang	.35
89.	Achterdochtig	.34
11.	Klampt zich vast aan volwassenen of is te afhankelijk	.33
14.	Huilt veel	.32
75.	Verlegen, schuchter	.30

4. HYPERACTIVE (HYPERACTIEF)

1.	Gedraagt zich te jong voor haar leeftijd	.65
62.	Onhandig, slechte coordinatie	.65
61.	Slechte schoolresultaten	.58
8.	Kan zich niet concentreren, kan niet lang de aandacht bij iets houden	.55
41.	Impulsief of onnadenkend gedrag	.48
64.	Speelt het liefst met jongere kinderen	.41
13.	In de war of chaotisch in het denken	.40
38.	Wordt veel geplaagd	.38
10.	Kan niet stil zitten, onrustig, of overactief	.38
79.	Spraakproblemen	.35
23.	Is ongehoorzaam op school	.33
102.	Traag,langzaam, te weinig energie	.33
17.	Dagdromen of gaat geheel op in haar gedachten	.31

5. SCHIZOID-OBSESSIVE (SCHIZOIDE - OBSESSIEF)

47.	Nachtmerries	.58
100.	Slaapmoeilijkheden	.58
70.	Ziet dingen die er niet zijn	.57
40.	Hoort dingen die er niet zijn	.53
29.	Is bang voor bepaalde dieren, situaties, of plaatsen, uitgezonderd de school	.50
50.	Is te angstig of te bang	.50
76.	Slaapt minder dan de meeste kinderen	.44
85.	Vreemde gedachten	.35
92.	Slaapwandelen of hardop praten in de slaap	.34
9.	Kan bepaalde gedachten niet uit zijn/haar hoofd zetten; obsessies	.33
54.	Oververmoeid	.33
31.	Is bang dat zij iets ondeugends of slechts zou kunnen doen of denken	.31

6. SOCIALLY WITHDRAWN (SOCIALE TERUGTREKKING)

111.	Teruggetrokken, komt niet tot contact met anderen	.68
69.	Gesloten, anderen weten niet wat er in haar omgaat	.61
80.	Kijkt met lege of 'wezenloze' blik	.52
75.	Verlegen, schuchter	.49
65.	Weigert om te praten	.49
102.	Traag, langzaam, te weinig 'energie'	.46
17.	Dagdromen of gaat geheel op in haar gedachten	.42
84.	Vreemd gedrag, doet vreemd aan (geef aan)	.36
42.	Vindt het fijn om alleen te zijn	.36

7. SOMATIC COMPLAINTS (SOMATISCHE KLACHTEN)
(lichamelijke klachten zonder duidelijke medische oorzaak)

56b.	Hoofdpijn	.68
56c.	Misselijkheid	.65
51.	Last van duizeligheid	.62
56f.	Maagpijn, buikpijn- of krampen	.52
56a.	Pijnen	.47
56d.	Moeilijkheden met zien	.41
77.	Slaapt meer dan de meeste kinderen overdag en/of 's nachts	.31

Syndromes of the boys aged 6-11

1. AGGRESSIVE (AGRESSIEF)

3.	Spreekt veel tegen of maakt veel ruzie	.74
104.	Is erg luidruchtig	.72
22.	Is thuis ongehoorzaam	.71
94.	Plaagt veel	.68
95.	Driftbuien of snel driftig	.67
86.	Koppig, stuurs of prikkelbaar	.66
41.	Impulsief of onnadenkend gedrag	.65
87.	Verandert plotseling van stemming	.65
19.	Eist veel aandacht op	.64
10.	Kan niet stilzitten, onrustig of overactief	.59
68.	Schreeuwt of gilt veel	.59
37.	Vecht veel	.57
74.	Speelt de clown, doet raar of gek om de aandacht te trekken	.53
88.	Mokken pruilen	.53
23.	Is ongehoorzaam op school	.53
7.	Opscheppen, stoer doen	.52
16.	Wreed, pesterig of gemeen voor anderen	.52
90.	Vloeken, schuttingtaal	.51
26.	Lijkt zich niet schuldig te voelen na zich misdragen te hebben	.49
27.	Snel jaloers	.48
20.	Vernielt eigen spullen	.46
93.	Praat te veel	.46
57.	Valt anderen aan of lastig	.41
8.	Kan zich niet concentreren, kan niet land de aandacht bij iets houden	.40

2. DELINQUENT (DELINQUENT)

81.	Steelt van huis	.71
82.	Steelt buitenshuis	.70
106.	Vandalisme, vernielen	.66
21.	Vernielt spullen van anderen gezinsleden of van andere kinderen	.62
43.	Liegen of bedriegen	.58
72.	Brandstichten	.57
20.	Vernielt eigen spullen	.52
15.	Wreed voor dieren	.47
97.	Bedreigt andere mensen	.35
39.	Gaat om met kinderen die in moeilijkheden verzeild raken	.32
16.	Wreed, pesterig of gemeen voor anderen	.32
23.	Is ongehoorzaam op school	.31

3. DEPRESSIVE (DEPRESSIEF)

33.	Klaagt erover of heeft het gevoel dat niemand van hem houdt	.64
103.	Ongelukkig, verdrietig, gedeprimeerd	.63
35.	Voelt zich waardeloos of minderwaardig	.63
34.	Heeft het gevoel dat anderen het op hem gemunt hebben	.63
12.	Klaagt over zich eenzaam voelen of in de steek gelaten voelen	.62
112.	Maakt zich zorgen	.47
30.	Is bang om naar school te gaan	.46
91.	Praat erover dat hij zichzelf zou willen doden	.42
89.	Achterdochtig	.38
52.	Te veel last van schuldgevoel	.36
9.	Kan bepaalde gedachten niet uit zijn hoofd zetten;obsessies	.31

4 HYPERACTIVE (HYPERACTIEF)

61.	Slechte schoolresultaten	.54
8.	Kan zich niet concentreren, kan niet lang de aandacht bij iets houden	.50

17.	Dagdrômen of gaat geheel op in zijn gedachten	.50
13.	In de war of chaotisch in het denken	.48
102.	Traag, langzaam, te weinig energie	.47
80.	Kijkt met lege of wezenloze blik	.47
1.	Gedraagt zich te jong voor zijn leeftijd	.46
62.	Onhandige, slechte coordinatie	.44
77.	Slaapt meer dan de meeste kinderen overdat en/of 's nachts	.35
79.	Spraakproblemen	.31

5 SOCIALLY WITHDRAWN (SOCIALE TERUGTREKKING)

25.	Kan niet goed opschieten met andere kinderen	.62
48.	Andere kinderen mogen hem niet	.61
38.	Wordt veel geplaagd	.57
64.	Speelt het liefst met jongere kinderen	.47
42.	Vindt het fijn om alleen te zijn	.43
111.	Teruggetrokken, komt niet tot contact met anderen	.43
34.	Heeft het gevoel dat anderen het om hem gemunt hebben	.32
84.	Vreemd gedrag, doet vreemd aan	.30

6. SOMATIC COMPLAINTS (SOMATISCHE KLACHTEN)
(lichamelijke klachten zonder medische oorzaak)

56.c	Misselijkheid	.76
56.a	Pijnen	.67
56.b	Hoofdpijn	.66
56.f	Maagpijn, buikpijn, of krampen	.63
51.	Last van duizeligheid	.61
56.g	Overgeven	.51
54.	Oververmoeidheid	.38

7. NON-COMMUNICATIVE (NIET-COMMUNICATIEF)

75.	Verlegen, schuchter	.58
69.	Gesloten, andere weten niet goed wat er in hem omgaat	.55
65.	Weigert om te praten	.53
71.	Schaamt of geneert zich sne	.42
111.	Teruggetrokken, komt niet tot contact met anderen	.41
99.	Overdreven netjes of schoon	.39
80.	Kijkt met lege of wezenloze blik	.38
42.	Vindt het fijn om alleen te zijn	.37
92.	Slaapwandelen of hardop praten in de slaap	.34
89.	Achterdochtig	.31
88.	Mokken, pruilen	.31

Appendix B

Example of the average classification score and the average typicality score for each item into the girls' syndrome 'Aggressive" .

Syndrome 1 girls: aggressive

	average classification score		the average typicality score	
	expertgroup	novicegroup	expertgroup	novicegroup
items	n=5	n=5	n=6	n=6
1			3.3	2.5
2	.6	.2	4.1	5.0
3	.6	.4	3.1	5.1
4	.8	1.0	5.8	7.0
5	.4	.8	5.5	5.0
6	.4	.2	4.5	3.1
7			3.0	1.8
8	.4	.6	4.3	4.6
9	1.0	.8	6.6	6.8
10	.4		5.8	5.3
11	.2		4.3	3.5
12			2.0	.3
13	.4	.4	5.0	6.1
14			.6	2.0
15			3.0	3.3
16		.2	3.5	4.3
17	.6	1.0	6.1	6.8
18	.8	1.0	6.0	6.0
19			.5	.3
20			5.6	5.3
21	.4	.2	5.8	5.5
22		.2	2.0	2.0
23			2.8	2.0
24	.6	.4	6.1	4.6
25	.4		1.6	1.5
26	.6	.4	4.0	3.8
27	.2	1.0	6.8	7.1
28			3.5*	1.0*
29			1.3	.6
30	.4		1.0	1.0
31	.2		3.5	1.1
32		.2	1.8	1.0
33	.8	1.0	6.6	6.6
34	.6	.4	.8	1.6
35			1.5	.3
36	.2		4.0	4.6
37	.2	.4	6.0*	3.5*
38			1.5	1.6
39	.2		2.1	1.3
40		.2	3.8	4.5
41	1.0	1.0	6.1	4.6
42			1.0	.5
43			3.3	.1
44	.4	.6	5.3	4.0
45	.2	.2	3.0	2.1
46	.4	.2	5.5	5.5
47			4.1	2.8

** The items for which the typicality ratings significantly differ between the expert versus the novice group are marked with an asterisk (*=p<.05)*

Appendix C

Core syndromes copied from Koot & Verhulst (1990)* .

Kernsyndroom 1: Agressief (Aggressive)

3 ruzie maken
7 opscheppen
16 pesterig
19 aandacht opeisen
22 thuis ongehoorzaam
26 voelt zich niet schuldig
27 snel jaloers
41 impulsief
104 luidruchtig
68 schreeuwen
74 gek doen
37 vecht veel
86 koppig prikkelbaar
67 stemmingswisselingen
88 mokken
90 vloeken, schelden
93 praat te veel
94 plaagt veel
95 driftbuien

Kernsyndroom 2: Angstig depressief (Anxious Depressed)

29 bang angstig
31 bang iets verkeerds te doen
30 bang voor school
33 voelt zich onbemind
52 te veel schuldgevoel
35 voelt zich minderwaardig
34 voelt zich achtervolgd
12 eenzaam
32 wil perfect zijn
45 zenuwachtig, gespannen
71 schaamt zich snel
103 ongelukkig, verdrietig
112 maakt zich zorgen

Kernsyndroom 3: Aandachtsproblemen (Attention Problems)

1 gedraagt zich te jong
8 kan zich niet concentreren
10 onrustig, overactief
13 in de war, chaotisch
17 dagdromen
41 impulsief
61 slechte schoolresultaten
62 onhandig, slechte coördinatie
80 lege blik

** The numbers refer to the number of the CBCL-items and the bolded items are used in the vignettes of Task 3.*

Kernsyndroom 4: Delinquent (Delinquent)

105 alcohol, drugs
39 slechte vrienden
43 liegen, bedriegen
21 vernielt spullen van anderen
23 ongehoorzaam op school
67 loopt weg van huis
72 brandstichten
81 steelt van huis
82 steelt buitenshuis
101 spijbelen
106 vandalisme

Kernsyndroom 5: Schizoïd (Schizoïd)

40 hoort geluiden, stemmen
66 dwanghandelingen
85 vreemde gedachten
70 ziet dingen
84 vreemd gedrag

Kernsyndroom 6: Lichamelijke klachten (Physical complaints)

56a pijnen
51 duizeligheid
56b hoofdpijn
56c misselijkheid
56f maagpijn, buikpijn
56g overgeven

Kernsyndroom 7: Teruggetrokken (Withdrawn)

42 liever alleen
65 weigert te praten
69 gesloten
71 schaamt zich snel
75 verlegen schuchter
80 lege blik
102 te weinig actief
103 ongelukkig verdrietig
11 teruggetrokken

Restgroepen

Alleen voor jongens:
8 sociaal onhandig (socially inept)

25 niet opschieten met andere kinderen
38 wordt geplaagd
48 niet geliefd
64 speelt lieftst met jongere kinderen

Alleen voor meisjes:
9 gemeen (cruel)

16 pesterig, gemeen
21 vernielt spullen van anderen
20 vernielt eigen spullen
57 valt anderen aan
37 vecht veel

Appendix D

Items omitted from the original CBCL item pool by the diagnosticians in Task 2.*

Diagnostician										
1	2	3	4	5	6	7	8	9	10	
								2	2	
								4		
	5	5	5							
	12	12								
14										
15		15								
18			18	18						
28	28		28		28		28	28		
29			29							
30		30	30							
31		31	31							
32										
		34								
			36							
39										
		40								
42										
51		51						51	51	
					52					
54			54							
57										
					58			58		
					59					
			60					60		
65			65							
67		67								
			68							
		70	70				70			
71										
		72								
		73					73			
		77								
		81								
			82							
		83			83		83			
						85				
					89					
	91	91								
					92			92		
					93					
		96								
97										
					98					
		99								
101	101									
					104					
105	105	105	105		105				105	
	110	110								
Total number of omissions										
19	5	19	17	1	11	1	4	7	3	87

* *The numbers are CBCL item numbers*

Appendix E

Items added to the CBCL item pool by diagnosticians in Task 2

Diagnostician 1

1. oninvoelbaar gedrag
2. impulsief
3. gedraagt zich ouder
4. star
5. geringe distantie tot volwassenen
6. egocentrisme
7. regressie, terugval naar jongere leeftijd.

Diagnostician 2

geen gedragingen toegevoegd

Diagnostician 3

1. ontwikkelingsretardatie
2. gehoorproblemen
3. fysische problemen
4. motorische problemen
5. mondmotoriek
6. oppositioneel gedrag/ verzet zich snel

Diagnostician 4

geen gedragingen toegevoegd

Diagnostician 5

1. heimwee
2. slaafs gedrag
3. aggressie gericht op bepaalde sexe
4. handtastelijk gedrag

Diagnostician 6

1. broekpoepen
2. gokverslaving
3. eigen aggressie structureren

Diagnostician 7

1. paniekreactie
2. depersonalisatie
3. echolalie
4. macropsie/ micropsie
5. gokverslaving
6. separatieproblemen
7. coprolalie
8. keelgeluiden, kuchen
9. lichamelijk handicap
10. epilepsie/ absences
11. lichamelijke ziekten

Diagnostician 8

1. klachten waarvan vermoed wordt dat ze een organisch substraat hebben, te weten, cerebraal tekort,dan wel dysfunctie.
2. klachten waarvan vaststaat dat er een cerebraal tekort of dysfunctie aan ten grondslag ligt en een evaluatie van de consequenties voor het neuropsychologisch functioneren wordt gevraagd.

Diagnostician 9

1. conformisme
2. dominant
3. narcisme

Diagnostician 10

geen gedragingen toegevoegd

Appendix F

Overview of intuitive clusters, including the cluster labels.

Diagnostician 1

1. neurotisch
2. hyperactiviteit
3. stoornis in contactname
4. depressief gedrag

Diagnostician 2

1. concreet gedrag:
   a. somatisch
   b. agressie naar de ander
   c. agressie op zichzelf gericht
   d. slaapproblemen
   e. psychotisch / bizar
   f. zindelijkheidsproblemen
   g bizar gedrag
   h. restgroep (concreet)
2 relationeel
3 negatief gewaardeerd / normatief
4. neurotisch gedrag
5. rest

Diagnostician 3

1. emotionele problemen
2. verzet gedrag (agressie a g.v. opvoeding)
3. oppositioneel gedrag (agressie vanuit kind zelf)
4. ontwikkelingsretardatie
5. functiestoornissen / motoriek
6. spraak-, taalproblemen
7. somatische problemen
8. bizar gedrag / psychiatrisch
9 stoornis in sexuele ontwikkeling
10. (niet gelabeld)
11. restgroep

Diagnostician 4

1. neurotisch / depressief
2. acting-out gedrag
3 impulsief
4. psychotisch
5. kleinkinderlijk gedrag
6. contact
7. functiestoornis
8. lage zelfwaardering
9. stoornis in sexualiteit
10. gedrag
11 eetproblemen
12 restgroep

Diagnostician 5

1. sociale kwetsbaarheid
2. agressie
3. milieuproblematiek
4. hersenbeschadiging / hyperactiviteit

5. vreemd gedrag
6. angstig
7. sexueel getint gedrag
8. neurotisch / dwangmatig
9. normoverschrijdend
10. somatisch
11. slaapproblemen

Diagnostician 6

1. geparentificeerd gedrag
2. verwaarloosd
3. te snel, te veel autonomie
4. te weinig autonomie
5. (geblokkeerd in) agressie
6. eetproblemen
7. sexuele problemen
8. identiteitsproblemen
9. counterfobisch / grote mate van onzekerheid
10. dwangmatig
11. mishandeld / angst voor de ouder
12. a-vitale kinderen / autistiform
13. neurologisch ongaaf / ADD-achtig
14. loyaliteitsproblemen / stabilisator van spanning bij e ouders
15. psychotiform / wanen
16. rest

Diagnostician 7

1. obsessioneel / neurotisch
2. psychotici
3. gedragsproblemen
4. oppositioneel
5. zwakzinnigen
6. infantielen / afhankelijken
7. agressie-regulatie problemen
8. ADHD
9. allergie
10. Gilles de la Tourette / licht-organisch
11. lichamelijke handicaps
12. hechtingsproblematiek
13. lichamelijke ziektes
14. te dik
15. sexuele identiteitsproblemen
16. kind dat niets mankeert
17. rest

Diagnostician 8

1. mutisme
2. depressief / sociaal-geïsoleerd
3. ADHD
4. acting-out / psychopathiforme pubers (is verlengde fase-problematiek van groep 3)
5. organisch trage kinderen
6. dwangmatig / obsessief-compulsief
7. zwakbegaafden

Diagnostician 9

1. opofferend / actief-afhankelijken
2. zich aanbiedend / psychosomatosen
3. dominant / passief gereserveerden
4. autonoom

5. afstandelijk / acting-out
6. terughoudend
7. onderdanig / ambivalent
8. hulpeloos / passief-afhankelijk
9. rest

Diagnostician 10

1. depressief
2. identiteitsprobleem
3. enuresis
4. ontwikkelingsretardatie
5. psychopathie
6. psychosomatiek
7. slaapmoeilijkheden
8. ontwikkelingsachterstand en pedagogische verwaarlozing
9. sexueel excessief gedrag
10. borderline
11. neurotisch
12. achterdochtige persoonlijkheid
13. dwangmatig
14. praakproblemen
15. encopresis
16. object-relatie problemen
17. ontwikkelingspsychotici
18. regulatiestoornis
19. eetstoornissen
20. secundaire milieu problemen
21. (niet gelabeld)
22. vloeken /schuttingtaal

Appendix G

Overview of the percentages expressing the similarity between the intuitive clusters and the core syndromes.*

Diagnostician 1

	Core syndrome number *externalizing*		*mixed*			*internalizing*		*boys only*	*girls only*	
	1	4	3	5	6	2	7	8	9	rest
number of items	(19)	(11)	(9)	(5)	(2)	(13)	(9)	(4)	(5)	
intuitive cluster 1 (32)	18.8%	6.3%	6.3%			25%	3.1%	9.4%	3.1%	28.1%
2 (29)	24.1%	10.3%	17.2%	3.4%	3.4%	3.4%	6.9%		10.3%	20.7%
3 (21)	13.8%			9.5%			9.5%			57.1%
4(22)	9.1%	9.1%	9.1%	9.1%			4.5%	9.1%		50%

Diagnostician 2

	Core syndrome number *externalizing*		*mixed*			*internalizing*		*boys only*	*girls only*	
	1	4	3	5	6	2	7	8	9	rest
number of items	(19)	(11)	(9)	(5)	(2)	(13)	(9)	(4)	(5)	
intuitive cluster 1 (17)	5.9%					52.9%	23.5%			17.6%
2 (17)	35.3%		23.5%				5.9%			35.3%
3(22)	31.8%	9.1%				9.1%	9.1%		13.6%	27.3%
4 (51)	5.9%	13.7%	9.8%	5.9%	3.9%	3.9%	2%		3.9%	51%
5(7)	42.9%				28.6%		14.3%			14.3%

** The percentage expresses how many items of a core syndrome are found in each intuitive cluster: 1 = aggressive; 2 = anxious depressed; 3 = attention problems ;4 = delinquent; 5 = schizoid; 6 = physical complaints; 7 = withdrawn; 8 = socially inept (boys only); 9 = cruel (girls only).*

Diagnostician 3

	Core syndrome number *externalizing*		*mixed*			*internalizing*		*boys only*	*girls only*	
	1	4	3	5	6	2	7	8	9	rest
number of items	(19)	(11)	(9)	(5)	(2)	(13)	(9)	(4)	(5)	
intuitive cluster 1 (51)	11.8%		9.8%	3.9%	2%	15.7%	11.8%	3.9%	3.9%	37.3%
2(35)	42.8%	17.1%	8.6%					2.8%	11.4%	17.1%
3(33)	42.4%	12.1%	12.1%	3.0%				6.1%	12.1%	12.1%
4(30)	6.7%		23.3%			6.7%	26.7%	3.3%		33.3%
5(8)			25%				12.5%			62.5%
6(10)			20%			20%	40%			20%
7(8)							12.5%			87.5%
8(9)			11.1%	11.1%			33.3%			44.4%
9(3)										100%
10(4)	25%			50%						25%
11(5)						40%				60%

Diagnostician 4

	Core syndrome number *externalizing*		*mixed*			*internalizing*		*boys only*	*girls only*	
	1	4	3	5	6	2	7	8	9	rest
number of items	(19)	(11)	(9)	(5)	(2)	(13)	(9)	(4)	(5)	
intuitive cluster 1 (36)			13.9%	2.8%	5.6%	13.9%	8.3%			55.6%
2(26)	30.8%	34.6%	3.8%			3.8%	3.8%		11.5%	11.5%
3(18)	27.8%		33.3%				5.5%		5.5%	27.8%
4(12)			25%	33.3%			8.3%			33.3%
5(13)	15.4%		7.7%							76.9%
6(9)	11.1%		11.1%					44.4%	11.1%	22.2%
7(8)			25%		12.5%		12.5%			50%
8(8)	12.5%		12.5%			37.5%	25%			12.5%
9(6)			16.7%							83.3%
10(6)	16.7%		16.7%							66.7%
11(3)										100%
12(4)	25%					25%	25%			25%

Diagnostician 5

	Core syndrome number *externalizing*		*mixed*			*internalizing*		*boys only*	*girls only*	
	1	4	3	5	6	2	7	8	9	rest
number of items	(19)	(11)	(9)	(5)	(2)	(13)	(9)	(4)	(5)	
intuitive cluster 1 (20	10%		10%			15%	25%	15%		25%
2(18)	50%	22.2%							5.6%	22.2%
3(14)	7.2%	28.6%				14.3%	7.2%		14.3%	28.6%
4(11)	18.2%		54.5%				9.1%			18.2%
5(10)	10%		10%	40%			10%			30%
6(10)						50%	10%	10%		30%
7(6)	16.7%									83.3%
8(7)				14.3%		42.9%				42.9%
9(9)	33.3%	33.3%							22.2%	11.1%
10(8)					25%					75%
11(6)										100%

Diagnostician 6

	Core syndrome number *externalizing*		*mixed*			*internalizing*		*boys only*	*girls only*	
	1	4	3	5	6	2	7	8	9	rest
number of items	(19)	(11)	(9)	(5)	(2)	(13)	(9)	(4)	(5)	
intuitive cluster 1 (4)			25%			25%				50%
2(15)	26.7%	40%							20%	13.3%
3(4)	25%	25%				25%				25%
4(8)	12.5%	12.5%	25%				12.5%	12.5%		25%
5(23)	4.3%	4.3%			8.7%	26.1%	13%		4.3%	39.1%
6(3)										100%
7(5)					20%	20%	20%			40%
8(2)										100%
9(10)	50%	20%							20%	10%
10(4)	25%			25%		25%				25%
11(5)										100%
12(9)			33.3				44.4%			22.2%
13(15)	13.3%		33.3%			6.7%		26.7%		20%
14(3)						33.3%	33.3%			33.3%
15(6)	16.7%		16.7%	66.7%						
16(1)	100%									

Diagnostician 7

	Core syndrome number externalizing		mixed			internalizing only		boys only	girls	
	1	4	3	5	6	2	7	8	9	rest
number of items	(19)	(11)	(9)	(5)	(2)	(13)	(9)	(4)	(5)	
intuitive cluster 1 (26)	3.5%		7.7%	7.7%	7.7%	42.3%	11.5%	3.8%		15.4%
2(28)	3.6%	3.6%	10.7%	14.3%		7.1%	10.7%	7.1%	3.6%	39.3%
3(31)	29%	35.5%	6.5%					6.5%	12.9%	9.7%
4(13)	53.8%	23.1%	7.7%					7.7%	7.7%	
5(10)	10%		10%					20%		60%
6(37)	13.5%	2.7%	16.2%		2.7%	10.8%	5.4%	5.4%		43.2%
7(10)	10%		10%					20%		60%
8(23)	21.7%	8.7%	21.7%			8.7%			4.3%	34.8%
9(6)	16.7%						16.7%			66.7%
10(5)			20%							80%
11(1)			100%							
12(16)	18.8%		12.5%			25%			6.3%	37.5%
13(1)			100%							
14(2)			50%							50%
15(3)			33.3%							66.7%
16(1)		100%								
17(1)							100%			

Diagnostician 8

	Core syndrome number externalizing		mixed			internalizing		boys only	girls only	
	1	4	3	5	6	2	7	8	9	rest
number of items	(19)	(11)	(9)	(5)	(2)	(13)	(9)	(4)	(5)	
intuitive cluster 1 (1)							100%			
2(43)		2.3%	9.3%		4.7%	27.9%	14%	7%	2.4%	23.6
3(33)	33.3%	30.3%	3%			6.1%	3%	3%	12.1%	9.1%
4(57)	31.6%	10.5%	10.5%			1.8%		5.3%	8.8%	31.6%
5(18)			16.7%	5.6%		11.1%	27.8%	22.2%		16.7%
6(43)		2.3%	14%	9.3%	2.3%	23.2%	11.6%	7%		30.2%
7(34)	11.8%	2.9%	11.8%	3%	3%	11.8%	2.9%	11.8%	5.9%	35.3%

Diagnostician 9

	Core syndrome number *externalizing*		*mixed*			*internalizing*		*boys only*	*girls only*	
	1	4	3	5	6	2	7	8	9	rest
number of items	(19)	(11)	(9)	(5)	(2)	(13)	(9)	(4)	(5)	
intuitive cluster 1 (9)	11.1%		22.2%			33.3%	11.1%			22.2%
2(17)	5.9%		17.6%		5.9%	5.9%				64.7%
3(12)		8.3%	16.7%	16.7%			25%	8.3%		25%
4(13)	61.5%		15.4%					15.4%		7.7%
5(22)	27.3%	31.8%	4.5%					4.5%	18.2%	13.6%
6(14)	14.3%	14.3%	14.3%			7.1%	21.4%	7.1%		21.4%
7(16)	6.3%	12.5%	6.3%					6.3%	6.3%	62.5%
8(24)			4.2%	12.5%		45.8%	12.5%	8.3%		16.7%
9(3)	33.3%		33.3%					33.3%		

Diagnostician 10

	Core syndrome number externalizing		mixed		internalizing		boys only	girls only		
number of items	1 (19)	4 (11)	3 (9)	5 (5)	6 (2)	2 (13)	7 (9)	8 (4)	9 (5)	rest
intuitive cluster 1 (23)			4.3%			30.4%	26.1%	13%		26%
2(5)								60%		40%
3(6)								50%		50%
4(7)							14.3%	42.9%		42.9%
5(19)	26.3%	36.8%						15.8%	10.5%	10.5%
6(6)		16.7%			16.7%	16.7%		50%		16.7%
7(7)						14.3%		42.9%		42.9%
8(21)	47.6%	9.5%					4.8%	19%		9.5%
9(7)	14.3%		14.3%					42.9%		28.6%
10(13)	30.8%		23.1%	15.4%				23.1%		28.6%
11(16)	25%	18.8%	6.3%			18.8%		18.8%	12.5%	
12(8)				12.5%		12.5%		37.5%		37.5%
13(14)				14.3%		28.6%	7.1%	21.4%		28.6%
14(1)										100%
15(7)	14.3%							42.8%		42.8%
16(9)	11.1%					11.1%	11.1%	33.3%		33.3%
17(13)	7.7%		15.4%	23.1%			15.4%	23.1%		15.4%
18(6)			33.3%					50%		16.7%
19(5)								40%		60%
20(2)			100%							
21(7)								42.9%		57.1%
22(8)	12.5%		25%					37.5%		25%

Curriculum Vitae

Nicole Krol, geboren op 28 november 1957, behaalde in 1976 het Atheneum diploma aan het Theresia lyceum te Tilburg. Een jaar later begon zij met de studie pedagogiek aan de Katholieke Universiteit van Nijmegen.
In 1985 studeerde zij aan deze universiteit af in de hoofdrichting Orthopedagogiek met als specialisatie gedrags- en opvoedingsmoeilijkheden. Tijdens en na haar studie heeft zij zich verdiept in de diagnostische besluitvorming. Zij heeft onderzoek gedaan naar het diagnostische besluitvormingsproces van een multidisciplinair team.
Van 1987 tot 1991 was zij werkzaam als AIO bij vakgroep Orthopedagogiek van de K.U. Nijmegen, bij de sectie Psychodiagnostiek. In dit proefschrift is het onderzoek gerapporteerd wat zij als AIO heeft uitgevoerd in een project dat als doel de ontwikkeling van expertsystemen heeft.

NICI
P.O. Box 9104, 6500 HE Nijmegen
Tel.: 080 - 612620